```
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 209/213 (98%), Positives = 212/213 (99%)

Query: 1   MTKEYEVEDMSKVAIVTGAGQGIGFAIAKRLHADGFKIGVLDYNEETAQAAVDKLSPEDA 60
           +TK+YEVEDMSKVAIVTGAGQGIGFAIAKRLHADGFKIG+LDYNEETAQAAVDKLSPEDA
Sbjct: 1   LTKKYEVEDMSKVAIVTGAGQGIGFAIAKRLHADGFKIGILDYNEETAQAAVDKLSPEDA 60

Query: 61  VAVVADVSKRDQVFDAFQKVVDTFGDLNVVVNNAGVAPTTPLDTITEEQFEKAFAINVGG 120
           VAVVADVSKRDQVFDAFQKVVDTFGDLNVVVNNAGVAPTTPLDTITEEQFEKAFAINVGG
Sbjct: 61  VAVVADVSKRDQVFDAFQKVVDTFGDLNVVVNNAGVAPTTPLDTITEEQFEKAFAINVGG 120

Query: 121 TIWGSQAAQKHFRELGHGGKIINATSQAGCEGNPNLTVYGGTKFAVRGITQTLAKDLASE 180
           TIWGSQAAQKHFRELGHGGKIINATSQAGCEGNPNLTVYGGTKFAVRGITQTLAKDLASE
Sbjct: 121 TIWGSQAAQKHFRELGHGGKIINATSQAGCEGNPNLTVYGGTKFAVRGITQTLAKDLASE 180

Query: 181 GITVNAYAPGIVKTPMMFDIAHEVGKNAGKDDE 213
           GITVNAYAPGIVKTPMMF IAHEVGKNAGKDDE
Sbjct: 181 GITVNAYAPGIVKTPMMFAIAHEVGKNAGKDDE 213
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1773

A DNA sequence (GBSx1880) was identified in *S.agalactiae* <SEQ ID 5513> which encodes the amino acid sequence <SEQ ID 5514>. This protein is predicted to be ATP-dependent DNA helicase. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3735(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB38451 GB:L47709 22.4% identity with Escherichia coli
           DNA-damage inducible protein ...; putative [Bacillus subtilis]
 Identities = 132/461 (28%), Positives = 231/461 (49%), Gaps = 22/461 (4%)

Query: 21  RKYAVVDLEATGAGPNAS--IIQVGIVIIQGNKIIDSYETDVNPHESLDEHIVHLTGITD 78
           +++ V+D+E TG  P     IIQ+  V+I+  +I + +   +NP++S+   I  LTGI++
Sbjct: 4   QRFVVIDVETTGNSPKKGDKIIQIAAVVIENGQITERFSKYINPNKSIPAFIEQLTGISN 63

Query: 79  KQLAKAPDFGQVAHHIYQLIEDCIFVAHNVKFDANLLAEQLFLEGCELRTPRI-DTVELS 137
           + +     F  VA  ++QL++   FVAHN+ FD   +  +L   G +L    + DTVELS
Sbjct: 64  QMVENEQPFEAVAEEVFQLLDGAYFVAHNIHFDLGFVKYELHKAGFQLPDCEVLDTVELS 123

Query: 138 QVFYPCLEKYSLGALAESLNIELTDAHTAIADARATAQLFIKLKAKISSLPKEVLETILT 197
           ++ +P  E Y L  L+E L +     H A +DA  T  +F+++  K+  LP   L+ +
Sbjct: 124 RIVFPGFEGYKLTELSEELQLRHDQPHRADSDAEVTGLIFLEILEKLRQLPYPTLKQLRR 183

Query: 198 FADNLLFESYLLIEEAYQEADFVNPKEYYFWQGLVLKKEKAVGKPKKLSSDFQ------- 250
            + + + +   L++    E        Y  +    +++ +A+        +F
Sbjct: 184 LSQHFISDLTHLLDMFINENRHTEIPGYTRFSSFSVREPEAIDVRINEDENFSFEIESWE 243

Query: 251 ------VNMALLGMDARPKQVVFADLVKAHFNDQTTTFLEAQPGLGKTYGYLLP--LLDQ 302
                 ++  + G + R  Q++    V   F ++    +EA PG+GKT GYL+P  L  +
Sbjct: 244 AGNEKALSELMPGYEKRDGQMMMMREVADAFANREHALIEAPPGIGKTIGYLIPAALFAK 303
```

```
Query: 303 SQKQQIIVSVPTKILQDQIMAKEIKHIQELFHIPCHS--IKGPRNYLKLDAFYKSLQVQD 360
             K+ +I+S  + +LQ QI+ K++  +Q+LF  P  +  +KG  +YL L  F + L  +D
Sbjct: 304 KSKKPVIISTYSTLLQQQILTKDLPIVQDLFPFPVTAAILKGQSHYLCLYKFEQVLHEED 363

Query: 361 RNRLINRFKMQLLVWLTETTTGDLDEIKQKQRLESYFDQLKHDGE-VTQSSLFYDLDFWK 419
            N      K QLLVWLTET TGD+ E+      +  +D+L +D +   +S   + + F++
Sbjct: 364 DNYDAVLTKAQLLVWLTETNTGDVAELNLPSGGKLLWDRLAYDDDSYKRSRSEHVIGFYE 423

Query: 420 RSYDKVAQSQLVIINHAYFL-ERVQDDKDFAKGKVLVFDEA 459
           R+     +S LVI NH+  L +     K   +    + DEA
Sbjct: 424 RAKQIAMRSDLVITNHSLLLTDEGSHKKRLPESGTFIIDEA 464
 Identities = 63/195 (32%), Positives = 88/195 (44%), Gaps = 16/195 (8%)

Query: 629 KVWIDTSMPNILDLSPEQYAYEIAKRLQDIMTLKQPT-LVLLTSKQTMFMVSDYLDKWEI 687
           +V I   M +I D    ++  + A+ ++ +   KQP  LVL TS   +  V       E+
Sbjct: 720 QVMIPKEMKSIQDTGQPEFIQDTARYIELMAKEKQPKILVLFTSHDMLKKVHQ-----EL 774

Query: 688 KH---------LTQD-KNGLAYNVKKRFDRGESNLLLGTGSFWEGVDFVHRDRLIEVITR 737
           KH         L Q    G    + K F      +LLGT  FWEGVDF   +    +I R
Sbjct: 775 KHNMSASGIQLLAQGITGGSPGKLMKTFKTSNQAILLGTNHFWEGVDFPGDELTTVMIVR 834

Query: 738 LPFDTPKDYFIQKLSQSLTKEGKNFFYDYSLPMTVLKLKQALGRTTRREEQKSAVIILDS 797
           LPF +P         +   K+GKN F   SLP  VL  +Q +GR  R    K  +IILD
Sbjct: 835 LPFRSPDHPLHAAKCELARKKGKNPFQTVSLPEAVLTFRQGIGRLLRSAGDKGTIIILDR 894

Query: 798 RLVIKSYGQTIMHSL 812
           R+    YG+  + +L
Sbjct: 895 RIKTAGYGRLFLDAL 909
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5515> which encodes the amino acid sequence <SEQ ID 5516>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3735(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 500/835 (59%), Positives = 626/835 (74%), Gaps = 2/835 (0%)

Query: 1   MFCFIDIACYNRLTMTQKKLRKYAVVDLEATGAGPNASIIQVGIVIIQGNKIIDSYETDV 60
           MFCFIDIACYNRLTMTQKKLRKYAVVDLEATGAGPNASIIQVGIVIIQGNKIIDSYETDV
Sbjct: 1   MFCFIDIACYNRLTMTQKKLRKYAVVDLEATGAGPNASIIQVGIVIIQGNKIIDSYETDV 60

Query: 61  NPHESLDEHIVHLTGITDKQLAKAPDFGQVAHHIYQLIEDCIFVAHNVKFDANLLAEQLF 120
           NPHESLDEHIVHLTGITDKQLAKAPDFGQVAHHIYQLIEDCIFVAHNVKFDANLLAE LF
Sbjct: 61  NPHESLDEHIVHLTGITDKQLAKAPDFGQVAHHIYQLIEDCIFVAHNVKFDANLLAEALF 120

Query: 121 LEGCELRTPRIDTVELSQVFYPCLEKYSLGALAESLNIELTDAHTAIADARATAQLFIKL 180
           LEG EL  PR+DTVEL+Q+F+P  EKY+L  L+  LNI+L +AHTAIADARATA LF++L
Sbjct: 121 LEGYELTIPRVDTVELAQLFFPRFEKYNLSHLSRQLNIDLAEAHTAIADARATAILFLRL 180

Query: 181 KAKISSLPKEVLETILTFADNLLFESYLLIEEAYQEADFVNPKEYYFWQGLVLKKEKAVG 240
             KI SLP E LE++L ++D+LLFE+ ++I+E   +A   +P +Y   + ++L K
Sbjct: 181 LQKIESLPIECLESLLVYSDSLLFETAMVIQEGLAKAKPYDPNKYIKIRQILLPKGSKAL 240

Query: 241 KPKKLSSDFQVNMALLGMDARPKQVVFADLVKAHFNDQTTTFLEAQPGLGKTYGYLLPLL 300
           KP ++S  F +NMALLG++ RPKQ  FA L+   ++    +F+EAQ G+GKTYGYLLPLL
Sbjct: 241 KPYQISKSFPINMALLGLEERPKQTQFAQLIDEDYHQGVASFIEAQTGIGKTYGYLLPLL 300

Query: 301 DQSQKQQIIVSVPTKILQDQIMAKEIKHIQELFHIPCHSIKGPRNYLKLDAFYKSLQVQD 360
            +  + QIIVSVPTK+LQDQ+MA E+  IQE FHI CHS+KGP NYLKLD+F  SL   D
```

```
Sbjct: 301 AKEDQNQIIVSVPTKLLQDQLMAGEVAAIQEQFHIACHSLKGPANYLKLDSFADSLDQND 360

Query: 361 RNRLINRFKMQLLVWLTETTTGDLDEIKQKQRLESYFDQLKHDGEVTQSSLFYDLDFWKR 420
           +NRL+NR+KMQLLVWL ET TGDLDEIKQKQR  +YF+QLKHDG++ QSS FYD DFW+
Sbjct: 361 QNRLVNRYKMQLLVWLLETKTGDLDEIKQKQRFAAYFEQLKHDGDIKQSSEFYDYDFWRV 420

Query: 421 SYDKVAQSQLVIINHAYFLERVQDDKDFAKGKVLVFDEAQKLVLGLENFSRGQLDISHQL 480
           SY+K   ++L+I NHAYFL RVQDDKDFA+ KVLVFDEAQKL+L L+  SR QL+++  L
Sbjct: 421 SYEKAKTARLLITNHAYFLHRVQDDKDFARNKVLVFDEAQKLMLQLDQLSRHQLNLTVFL 480

Query: 481 QVIQKIIDSSIPLLQKRLLESISYELSHAVELFYRHNSFEFSETWLKRLKNSINALEVVG 540
           Q IQ  + + +PLL+KRLLES+S+EL      +Y++   + +  W  R+      L
Sbjct: 481 QTIQAKLSNPLPLLEKRLLESLSFELGQVSSDYYQNKEHQLAHDW-SRIAGYAKELTGAD 539

Query: 541 LDELQTFFTATYTNYWFETDKVNEKRLTILRGAREDFLKFSKFLPPTKKTYMISATLQIS 600
             ELQ FF  +  +YW  ++K  EKR+T L  A + F+ F + LP T KTY +SATL IS
Sbjct: 540 YQELQAFFATSDGDYWLSSEKQEEKRVTYLNSASKAFIHFQQLLPETVKTYFVSATLTIS 599

Query: 601 PKVYLSDLLGGFSSISTEKIAHEKNANQKVWIDTSMPNILDLSPEQYAYEIAKRLQDIMT 660
            +V L+DLL GF       I  +K  +Q V +D   P + ++S + Y   IAKR++ +
Sbjct: 600 SEVTLADLL-GFEEYLYHVIEKDKKQDQLVLVDQEAPIVTEVSDQIYVEAIAKRIESLKQ 658

Query: 661 LKQPTLVLLTSKQTMFMVSDYLDKWEIKHLTQDKNGLAYNVKKRFDRGESNLLLGTGSFW 720
              P LVL  SK+ + +VSDYLD+W++ HL Q+KNG AYN+KKRFD+GE  +LLG GSFW
Sbjct: 659 EGYPILVLFNSKKHLLLVSDYLDQWQVPHLAQEKNGTAYNIKKRFDQGEQTILLGLGSFW 718

Query: 721 EGVDFVHRDRLIEVITRLPFDTPKDYFIQKLSQSLTKEGKNFFYDYSLPMTVLKLKQALG 780
           EGVDF+  DR+I +I RLPFD P+D+F++K+S  L ++GKN F DY LPMT+L+LKQA+G
Sbjct: 719 EGVDFIQADRMITLIARLPFDNPEDFFVKKMSHYLLEKGKNPFRDYFLPMTILRLKQAIG 778

Query: 781 RTTRREEQKSAVIILDSRLVIKSYGQTIMHSLGRDFEISKEKINKVLTEMAKFLI 835
           RT RR++QKS VIILD RL+ KSYGQ I+  LG++F IS++  +  L E   FLI
Sbjct: 779 RTMRRQDQKSVVIILDRRLLTKSYGQVILEGLGQEFLISQQNFHDCLVETDCFLI 833
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1774

A DNA sequence (GBSx1881) was identified in *S.agalactiae* <SEQ ID 5517> which encodes the amino acid sequence <SEQ ID 5518>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2042(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9633> which encodes amino acid sequence <SEQ ID 9634> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF12702 GB:AF035157 aspartate aminotransferase [Lactococcus
          lactis]
 Identities = 270/391 (69%), Positives = 314/391 (80%)

Query: 7    MTYLSERVLNMEESVTLAAGAKARELRVQGRDILSLTLGEPDFATPKNIQQAAIEAITDG 66
            M   S+ VL M+ESVTLAA  +A+ L+ QGRDI+ LTLG+PDF TPK I QAAIEAI +G
Sbjct: 1    MKKCSDFVLKMDESVTLAAANRAKALKAQGRDIIDLTLGQPDFPTPKKIGQAAIEAINNG 60

Query: 67   RASFYTPSSGLPELKSAINAYFERFYGYSLKPNQVVVGTGAKFILYTFFMTVLNPGDEVI 126
            +ASFYT + GLPELK A+  Y+ RFY Y ++ N++++  GAKF LY +FM  ++P DEVI
```

```
Sbjct: 61  QASFYTQAGGLPELKKAVQHYWTRFYAYEIQTNEILITAGAKFALYAYFMATVDPLDEVI 120

Query: 127 IPTPYWVSYADQIKMAEGKPVFVTAKEVNHFKVTVEQLEAVRTDKTKVILLNSPSNPTGM 186
           IP PYWVSY DQ+KMA G PV V AK+ N+FKVTVEQLE  RT KTK++LLNSPSNPTGM
Sbjct: 121 IPAPYWVSYVDQVKMAGGNPVIVEAKQENNFKVTVEQLEKARTSKTKILLLNSPSNPTGM 180

Query: 187 IYKAEELEAIGNWAVEHDILILADDIYGRLVYNGNIFTPISSLSESIRNQTIVINGVSKT 246
           IY  EEL AIG WAV HD+LILADDIY RLVYNG  FT ISSLS+ IRN+T VINGVSKT
Sbjct: 181 IYSKEELTAIGEWAVAHDLLILADDIYHRLVYNGAEFTAISSLSDEIRNRTTVINGVSKT 240

Query: 247 YAMTGWRVGFAVGNHDIIAAMSKVVSQTTSNLTAVSQYATIEALNGSQESFEKMRLAFEE 306
           +AMTGWR+G AVG+ +IIAAM+K+ SQTTSN TAV+QYA IEA   + +SFEKM  AFEE
Sbjct: 241 FAMTGWRIGLAVGDPEIIAAMTKIASQTTSNPTAVAQYAAIEAFEENDKSFEKMHAAFEE 300

Query: 307 RLNIIYPLLCQVPGFEVVKPQGAFYLFPNVTKAMEMKGYTDVTAFTDAILEEVGLALVTG 366
           RLN IY  L +VPGFE+VKP GAFYLFP VTKAM MKGYTDVT FT AILEE G+ALVTG
Sbjct: 301 RLNKIYLQLSEVPGFELVKPNGAFYLFPKVTKAMAMKGYTDVTDFTTAILEEAGVALVTG 360

Query: 367 AGFGAPENVRLSYATDLETLKEAVRRLHVFM 397
           AGFG+PENVRLSYAT LETL+ AV RL  +M
Sbjct: 361 AGFGSPENVRLSYATSLETLEAAVTRLKDWM 391
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1005> which encodes the amino acid sequence <SEQ ID 1006>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.48   Transmembrane   95 - 111 (  95 - 113)

----- Final Results -----
              bacterial membrane --- Certainty=0.1192(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 301/397 (75%), Positives = 343/397 (85%)

Query: 7   MTYLSERVLNMEESVTLAAGAKARELRVQGRDILSLTLGEPDFATPKNIQQAAIEAITDG 66
           M  LS+RVL M+ESVTLAAGA+A+ L+ QGRD+L+LTLGEPDF TPK+IQ  AIE+I +G
Sbjct: 1   MPKLSKRVLEMKESVTLAAGARAKALKAQGRDVLNLTLGEPDFFTPKHIQDKAIESIQNG 60

Query: 67  RASFYTPSSGLPELKSAINAYFERFYGYSLKPNQVVVGTGAKFILYTFFMTVLNPGDEVI 126
            ASFYT +SGLPELK+AI  Y +  YGY L P+Q+V GTGAKFILY FFM VLNPGD+V+
Sbjct: 61  TASFYTNASGLPELKAAIATYLKNQYGYHLSPDQIVAGTGAKFILYAFFMAVLNPGDQVL 120

Query: 127 IPTPYWVSYADQIKMAEGKPVFVTAKEVNHFKVTVEQLEAVRTDKTKVILLNSPSNPTGM 186
           IPTPYWVSY+DQ+KMAEG+P+FV   E N FKVTV+QLE  RT KTKV+L+NSPSNPTGM
Sbjct: 121 IPTPYWVSYSDQVKMAEGQPIFVQGLEENQFKVTVDQLERARTSKTKVVLINSPSNPTGM 180

Query: 187 IYKAEELEAIGNWAVEHDILILADDIYGRLVYNGNIFTPISSLSESIRNQTIVINGVSKT 246
           IY AEEL AIG WAV +DILILADDIYG LVYNGN F PIS+LSE+IR QTI +NGV+K+
Sbjct: 181 IYGAEELRAIGEWAVHNDILILADDIYGSLVYNGNQFVPISTLSEAIRRQTITVNGVAKS 240

Query: 247 YAMTGWRVGFAVGNHDIIAAMSKVVSQTTSNLTAVSQYATIEALNGSQESFEKMRLAFEE 306
           YAMTGWRVGFA G  +II+AMSK++ QTTSNLT VSQYA IEA  GSQ S E+MRLAFEE
Sbjct: 241 YAMTGWRVGFAAGEPEIISAMSKIIGQTTSNLTTVSQYAAIEAFCGSQSSLEEMRLAFEE 300

Query: 307 RLNIIYPLLCQVPGFEVVKPQGAFYLFPNVTKAMEMKGYTDVTAFTDAILEEVGLALVTG 366
           RLNI YPLLCQVPGFEVVKPQGAFY FPNV KAMEM G++DVT+F +AILEEVGLA+V+G
Sbjct: 301 RLNITYPLLCQVPGFEVVKPQGAFYFFPNVKKAMEMTGFSDVTSFANAILEEVGLAVVSG 360

Query: 367 AGFGAPENVRLSYATDLETLKEAVRRLHVFMGSNEIN 403
           AGFGAPENVRLSYATD+ETLKEAVRRLHVFM SNEIN
Sbjct: 361 AGFGAPENVRLSYATDIETLKEAVRRLHVFMESNEIN 397
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1775

A DNA sequence (GBSx1882) was identified in *S.agalactiae* <SEQ ID 5519> which encodes the amino acid sequence <SEQ ID 5520>. This protein is predicted to be asparaginyl-tRNA synthetase (asnS). Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1488(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB05415 GB:AP001512 asparaginyl-tRNA synthetase [Bacillus halodurans]
 Identities = 252/442 (57%), Positives = 316/442 (71%), Gaps = 15/442 (3%)

Query: 7   SIVDVKDYVGQEVTIGAWVANKSGKGKIAFVQLRDGSAFFQGVAFKPNFIEKYGEESGLE 66
           +I  + YV QEVT+GAW+ANK   GKIAF+QLRDG+ F QGV  K         E G E
Sbjct: 4   TIAKIGQYVDQEVTLGAWLANKRSSGKIAFLQLRDGTGFIQGVVVKA--------EVGDE 55

Query: 67  KFDVIKRLNQETSVYVTGIVKEDERSKFGYELDITDLEVIGESHEYPITPKEHGTDFLMD 126
            F   K L QE+S+YVTGIV++DER+  GYEL +T  ++I E+ +YPITPKEHGT+FLMD
Sbjct: 56  WFQKAKNLTQESSLYVTGIVRKDERAPSGYELTVTSFDIIHEATDYPITPKEHGTEFLMD 115

Query: 127 NRHLWLRSRKQMAVMQIRNAIIYSTYEFFDQNGFIKFDSPILSENAAEDSTELFETDYFG 186
           +RHLW+RSRKQ AV++IRN II +TYEFF +NGF+K D PIL+ +A E +TELF T YF
Sbjct: 116 HRHLWIRSRKQHAVLRIRNEIIRATYEFFHENGFVKVDPPILTGSAPEGTTELFHTKYFD 175

Query: 187 KPAFLSQSGQLYLEAGAMALGRVFDFGPVFRAEKSKTRRHLTEFWMMDAEYSFLSHEESL 246
           + AFLSQSGQLY+EA A+A GRVF FGP FRAEKSKTRRHL EFWM++ E +F+  EESL
Sbjct: 176 EDAFLSQSGQLYMEAAALAFGRVFSFGPTFRAEKSKTRRHLIEFWMIEPEMAFVEFEESL 235

Query: 247 DLQEAYVKALIQGVLDRAPQALDILERDVEALKRYIAEPFKRVSYDDAITLLQEHEADED 306
           ++QE YV  ++Q VL      L  L RD   L+  I  PF R+SYDDAI  L E   D+
Sbjct: 236 EIQENYVAYIVQSVLKHCAIELKTLGRDTSVLES-IQAPFPRISYDDAIKFLHEKGFDD- 293

Query: 307 TDYEHLEHGDDFGSPHETWISNYFGVPTFVVNYPASFKAFYMKPVPGNPERVLCADLLAP 366
                +E GDDFG+PHET I+ +F  P F+ +YP S K FYM+P P    + VLCADL+AP
Sbjct: 294 -----IEWGDDFGAPHETAIAEHFDKPVFITHYPTSLKPFYMEPDPNRDDVVLCADLIAP 348

Query: 367 EGYGEIIGGSMREDDYDALVAKMDELGMDKSEYDFYLDLRKYGSVPHGGFGIGIERMVTF 426
           EGYGEIIGGS R  DYD L  +++E  +    Y +YLDLRKYGSVPH GFG+G+ER V +
Sbjct: 349 EGYGEIIGGSQRISDYDLLKKRLEEHDLSLDAYAWYLDLRKYGSVPHSGFGLGLERTVGW 408

Query: 427 VAGTKHIREAIPFPRMLHRIKP 448
           ++G  H+RE IPFPR+L+R+ P
Sbjct: 409 ISGAGHVRETIPFPRLLNRLYP 430
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5521> which encodes the amino acid sequence <SEQ ID 5522>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1488(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 443/448 (98%), Positives = 447/448 (98%)

Query: 1   MSKKLISIVDVKDYVGQEVTIGAWVANKSGKGKIAFVQLRDGSAFFQGVAFKPNFIEKYG 60
           MSKKLISIVDVKDYVGQEVTIGAWVANKSGKGKIAFVQLRDGSAFFQGVAFKPNFIEKYG
Sbjct: 1   MSKKLISIVDVKDYVGQEVTIGAWVANKSGKGKIAFVQLRDGSAFFQGVAFKPNFIEKYG 60

Query: 61  EESGLEKFDVIKRLNQETSVYVTGIVKEDERSKFGYELDITDLEVIGESHEYPITPKEHG 120
           EESGLEKFDVIKRLNQETSVYVTGIVKEDERSKFGYELDITDLE+IGESHEYPITPKEHG
Sbjct: 61  EESGLEKFDVIKRLNQETSVYVTGIVKEDERSKFGYELDITDLEIIGESHEYPITPKEHG 120

Query: 121 TDFLMDNRHLWLRSRKQMAVMQIRNAIIYSTYEFFDQNGFIKFDSPILSENAAEDSTELF 180
           TDFLMDNRHLWLRSRKQMAVMQIRNAIIY+TYEFFDQNGFIKFDSPILSENAAEDSTELF
Sbjct: 121 TDFLMDNRHLWLRSRKQMAVMQIRNAIIYATYEFFDQNGFIKFDSPILSENAAEDSTELF 180

Query: 181 ETDYFGKPAFLSQSGQLYLEAGAMALGRVFDFGPVFRAEKSKTRRHLTEFWMMDAEYSFL 240
           ETDYFGKPAFLSQSGQLYLEAGAMALGRVFDFGPVFRAEKSKTRRHLTEFWMMDAEYSFL
Sbjct: 181 ETDYFGKPAFLSQSGQLYLEAGAMALGRVFDFGPVFRAEKSKTRRHLTEFWMMDAEYSFL 240

Query: 241 SHEESLDLQEAYVKALIQGVLDRAPQALDILERDVEALKRYIAEPFKRVSYDDAITLLQE 300
           SHEESLDLQEAYVKALIQGVLDRAPQALDILERDVEALKRYI EPFKRVSYDDAITLLQE
Sbjct: 241 SHEESLDLQEAYVKALIQGVLDRAPQALDILERDVEALKRYITEPFKRVSYDDAITLLQE 300

Query: 301 HEADEDTDYEHLEHGDDFGSPHETWISNYFGVPTFVVNYPASFKAFYMKPVPGNPERVLC 360
           HEADEDTDYEHLEHGDDFGSPHETWISNYFGVPTFVVNYPASFKAFYMKPVPGNPERVLC
Sbjct: 301 HEADEDTDYEHLEHGDDFGSPHETWISNYFGVPTFVVNYPASFKAFYMKPVPGNPERVLC 360

Query: 361 ADLLAPEGYGEIIGGSMREDDYDALVAKMDELGMDKSEYDFYLDLRKYGSVPHGGFGIGI 420
           ADLLAPEGYGEIIGGSMRED+YDALVAKMDELGMDKSEYDFYLDLRKYGSVPHGGFGIGI
Sbjct: 361 ADLLAPEGYGEIIGGSMREDNYDALVAKMDELGMDKSEYDFYLDLRKYGSVPHGGFGIGI 420

Query: 421 ERMVTFVAGTKHIREAIPFPRMLHRIKP 448
           ERMVTFVAGTKHIREAIPFPRMLHRI+P
Sbjct: 421 ERMVTFVAGTKHIREAIPFPRMLHRIRP 448
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1776

A DNA sequence (GBSx1883) was identified in *S.agalactiae* <SEQ ID 5523> which encodes the amino acid sequence <SEQ ID 5524>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -6.85   Transmembrane  103 - 119 ( 102 - 127)
   INTEGRAL    Likelihood = -5.04   Transmembrane   73 -  89 (  68 -  93)
   INTEGRAL    Likelihood = -4.19   Transmembrane   31 -  47 (  31 -  49)
   INTEGRAL    Likelihood = -1.86   Transmembrane  157 - 173 ( 157 - 173)

----- Final Results -----
               bacterial membrane --- Certainty=0.3739(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD40355 GB:AF036485 hypothetical protein [Plasmid pNZ4000]
 Identities = 39/135 (28%), Positives = 72/135 (52%), Gaps = 4/135 (2%)

Query: 3   KSPARLISFISIAIAINLVGANLALFLRLPIYLDTIGTLLIAVILGPWYAASTAFLSALI 62
           K  A  ++ I  A+ IN V   LA  L+LP++L ++GT L +++ GP   A + F++ +I
Sbjct: 15  KLSAATMTLIPAAVGINYVAKALAEGLKLPVWLGSLGTFLASMLAGPVAGAISGFINNVI 74

Query: 63  NWMTTDIFSLYYSPVAIVVAIITGILIKRNCKPSS--LLWKSLIISLPGTIIASVITVIL 120
```

```
             +T    S  Y+  +I + I  G+L       S+  +   ++II++   +I++ + VI
Sbjct: 75   YGLTLSPISTVYAITSIGIGIAVGVLHANGWFSSARRVFVSAIIIAIVSAVISTPLNVIF 134

Query: 121 FKGIT--SSGSSIIA 133
           + G T  + G S+ A
Sbjct: 135 WGGQTGIAWGDSLFA 149
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1777

A DNA sequence (GBSx1884) was identified in *S.agalactiae* <SEQ ID 5525> which encodes the amino acid sequence <SEQ ID 5526>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1873(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC75223 GB:AE000305 orf, hypothetical protein [Escherichia coli K12]
 Identities = 97/305 (31%), Positives = 160/305 (51%), Gaps = 10/305 (3%)

Query: 1   MNKEKIIIDCDPGIDDTLALMYAIQHPKLEVVAITITAGNSPVELGLKNTFVTLELLNRH 60
           M K KII+DCDPG DD +A+M A +HP ++++ ITI AGN  ++  L N     + L
Sbjct: 1   MEKRKIILDCDPGHDDAIAIMMAAKHPAIDLLGITIVAGNQTLDKTLINGLNVCQKL-EI 59

Query: 61  DIPVYVGDNLPLQREFVSAQDTHGMDGLGENNFTLAQPIIFQEESADC---FLANYFEHK 117
           ++PVY G   P+ R+ + A + HG  GL    F   +P+  Q ES       +
Sbjct: 60  NVPVYAGMPQPIMRQQIVADNIHGETGLDGPVF---EPLTRQAESTHAVKYIIDTLMASD 116

Query: 118 NDTSIIALGPLTNIARALQTNPKLGKHCKRFISMGGSFKSHGNCSPVAEYNYWCDPHAAQ 177
            D +++ +GPL+NIA A++  P +    +  + MGG++ + GN +P AE+N + DP AA+
Sbjct: 117 GDITLVPVGPLSNIAVAMRMQPAILPKIREIVLMGGAYGT-GNFTPSAEFNIFADPEAAR 175

Query: 178 YVFENLDKKIEMVGLDITRHIVLTPNHLSYMERINPDVSSFIQKITKFYFDFHWQYEHII 237
            VF +    + M+GLD+T   V TP+ ++ MER           I  F     ++   +
Sbjct: 176 VVFTS-GVPLVMMGLDLTNQTVCTPDVIARMERAGGPAGELFSDIMNFTLKTQFENYGLA 234

Query: 238 GCVINDPLAIAYFVNENIATGFDSYTDVACH-GIAMGQTIVDQYHFYKKDANSKILTSVN 296
           G  ++D   I Y +N +     + Y +V  + G   G+T+ D+     K AN+K+  +++
Sbjct: 235 GGPVHDATCIGYLINPDGIKTQEMYVEVDVNSGPCYGRTVCDELGVLGKPANTKVGITID 294

Query: 297 TNLFW 301
           T+ FW
Sbjct: 295 TDWFW 299
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1778

A DNA sequence (GBSx1885) was identified in *S.agalactiae* <SEQ ID 5527> which encodes the amino acid sequence <SEQ ID 5528>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1860(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB62728 GB:AL133423 hypothetical protein SC4A7.24c
          [Streptomyces coelicolor A3(2)]
 Identities = 36/134 (26%), Positives = 57/134 (41%), Gaps = 7/134 (5%)

Query: 1   MLYEVTSSNTQGVDGKVYLSNGKIVETNHPLNHL----PGFNPEELIALAWSTCLNATIK 56
           +LY   ++   G DG+V   +G++    +P   +     G NPE+L A  +S C    +
Sbjct: 8   VLYTAVATAENGRDGRVATDDGRLDVVVNPPKEMGGNGAGTNPEQLFAAGYSACFQGALG 67

Query: 57  AILEQKGFKDLKSRVDVTCQLMKEKQVGKGFYFQVNAVASIEKLSLSDSKLIVNKAHSRC 116
            +  Q+G     S V     + K      GF   V   A I  +  + ++ +V KAH  C
Sbjct: 68  VVARQEGADISGSTVTAKVGIGKNDD---GFGIIVEISAEIPTVDAATARSLVEKAHQVC 124

Query: 117 PISKLISNAKTINL 130
           P SK      T+ L
Sbjct: 125 PYSKATRGNITVTL 138
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1779

A DNA sequence (GBSx1886) was identified in *S.agalactiae* <SEQ ID 5529> which encodes the amino acid sequence <SEQ ID 5530>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0531(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9635> which encodes amino acid sequence <SEQ ID 9636> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15482 GB:Z99121 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 164/285 (57%), Positives = 207/285 (72%), Gaps = 2/285 (0%)

Query: 6   IKLVIVTGMSGAGKTVAIQSFEDLGYFTIDNMPPTLVPKFLELAAQSGDT-SKIAMVVDM 64
           I+LVI+TGMSGAGKTVAIQSFEDLGYF +DN+PP+L+PKFLEL  +S    SK+A+V+D+
Sbjct: 9   IQLVIITGMSGAGKTVAIQSFEDLGYFCVDNLPPSLLPKFLELMKESNSKMSKVALVMDL 68

Query: 65  RSRLFFREINSILDSLEINDNINFKILFLDATDTELVSRYKETRRSHPLAADGRVLDGIS 124
           R R FF  +   LD +  N  I  +ILFLDA D+ LV+RYKETRRSHPLAA G  L+GI+
Sbjct: 69  RGREFFDRLIEALDEMAENPWITPRILFLDAKDSILVTRYKETRRSHPLAATGLPLEGIA 128

Query: 125 LERELLAPLKSMSQNVVDTSELTPRQLRKVISKEFSNQDSQSSFRIEVMSFGFKYGIPLD 184
           LERELL  LK  SQ + DTS++ PR LR+ I K F+    ++ F + VMSFGFKYGIP+D
Sbjct: 129 LERELLEELKGRSQIIYDTSDMKPRDLREKIVKHFATNQGET-FTVNVMSFGFKYGIPID 187

Query: 185 ADLVFDVRFLPNPYYKPELRDKTGLDTEVYDYVMSFDESDDFYDHLLALIKPILPGYQNE 244
```

```
           ADLVFDVRFLPNPYY   +R  TG D EV  YVM ++E+  F + L+ L+  +LP Y+ E
Sbjct: 188 ADLVFDVRFLPNPYYIESMRPLTGKDKEVSSYVMKWNETQKFNEKLIDLLSFMLPSYKRE 247

Query: 245 GKSVLTVAIGCTGGQHRSTAFAHRLSEDLKADWTVNESHRDKNKR 289
           GKS + +AIGCTGGQHRS   A  L++  K D+  + +HRD  KR
Sbjct: 248 GKSQVVIAIGCTGGQHRSVTLAENLADYFKKDYYTHVTHRDIEKR 292
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5531> which encodes the amino acid sequence <SEQ ID 5532>. Analysis of this protein sequence reveals the following:

```
        Possible site: 20
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB15482 GB:Z99121 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 164/291 (56%), Positives = 213/291 (72%), Gaps = 3/291 (1%)

Query: 1   MSDKH-INLVIVTGMSGAGKTVAIQSFEDLGYFTIDNMPPALVPKFLELIEQTNENR-RV 58
           +S+ H I LVI+TGMSGAGKTVAIQSFEDLGYF +DN+PP+L+PKFLEL++++N    +V
Sbjct: 3   VSESHDIQLVIITGMSGAGKTVAIQSFEDLGYFCVDNLPPSLLPKFLELMKESNSKMSKV 62

Query: 59  ALVVDMRSRLFFKEINSTLDSIESNPSIDFRILFLDATDGELVSRYKETRRSHPLAADGR 118
           ALV+D+R R FF  +   LD +  NP I  RILFLDA D  LV+RYKETRRSHPLAA G
Sbjct: 63  ALVMDLRGREFFDRLIEALDEMAENPWITPRILFLDAKDSILVTRYKETRRSHPLAATGL 122

Query: 119 VLDGIRLERELLSPLKSMSQHVVDTTKLTPRQLRKTISDQFSEGSNQASFRIEVMSFGFK 178
            L+GI LERELL  LK  SQ + DT+ + PR LR+ I   F+    + +F + VMSFGFK
Sbjct: 123 PLEGIALERELLEELKGRSQIIYDTSDMKPRDLREKIVKHFATNQGE-TFTVNVMSFGFK 181

Query: 179 YGLPLDADLVFDVRFLPNPYYQVELREKTGLDEDVFNYVMSHPESEVFYKHLLNLIVPIL 238
           YG+P+DADLVFDVRFLPNPYY   +R  TG D++V +YVM   E++ F + L++L+  +L
Sbjct: 182 YGIPIDADLVFDVRFLPNPYYIESMRPLTGKDKEVSSYVMKWNETQKFNEKLIDLLSFML 241

Query: 239 PAYQKEGKSVLTVAIGCTGGQHRSVAFAHCLAESLATDWSVNESHRDQNRR 289
           P+Y++EGKS + +AIGCTGGQHRSV  A  LA+    D+  + +HRD  +R
Sbjct: 242 PSYKREGKSQVVIAIGCTGGQHRSVTLAENLADYFKKDYYTHVTHRDIEKR 292
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 234/296 (79%), Positives = 263/296 (88%)

Query: 1   MSDEQIKLVIVTGMSGAGKTVAIQSFEDLGYFTIDNMPPTLVPKFLELAAQSGDTSKIAM 60
           MSD+ I LVIVTGMSGAGKTVAIQSFEDLGYFTIDNMPP LVPKFLEL  Q+ +  ++A+
Sbjct: 1   MSDKHINLVIVTGMSGAGKTVAIQSFEDLGYFTIDNMPPALVPKFLELIEQTNENRRVAL 60

Query: 61  VVDMRSRLFFREINSILDSLEINDNINFKILFLDATDTELVSRYKETRRSHPLAADGRVL 120
           VVDMRSRLFF+EINS LDS+E N +I+F+ILFLDATD ELVSRYKETRRSHPLAADGRVL
Sbjct: 61  VVDMRSRLFFKEINSTLDSIESNPSIDFRILFLDATDGELVSRYKETRRSHPLAADGRVL 120

Query: 121 DGISLERELLAPLKSMSQNVVDTSELTPRQLRKVISKEFSNQDSQSSFRIEVMSFGFKYG 180
           DGI LERELL+PLKSMSQ+VVDT++LTPRQLRK IS +FS   +Q+SFRIEVMSFGFKYG
Sbjct: 121 DGIRLERELLSPLKSMSQHVVDTTKLTPRQLRKTISDQFSEGSNQASFRIEVMSFGFKYG 180

Query: 181 IPLDADLVFDVRFLPNPYYKPELRDKTGLDTEVYDYVMSFDESDDFYDHLLALIKPILPG 240
           +PLDADLVFDVRFLPNPYY+ ELR+KTGLD +V++YVMS  ES+ FY HLL LI PILP
Sbjct: 181 LPLDADLVFDVRFLPNPYYQVELREKTGLDEDVFNYVMSHPESEVFYKHLLNLIVPILPA 240

Query: 241 YQNEGKSVLTVAIGCTGGQHRSTAFAHRLSEDLKADWTVNESHRDKNKRKETVNRS 296
           YQ EGKSVLTVAIGCTGGQHRS AFAH L+E L  DW+VNESHRD+N+RKETVNRS
Sbjct: 241 YQKEGKSVLTVAIGCTGGQHRSVAFAHCLAESLATDWSVNESHRDQNRRKETVNRS 296
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1780

A DNA sequence (GBSx1887) was identified in *S.agalactiae* <SEQ ID 5533> which encodes the amino acid sequence <SEQ ID 5534>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB96620 GB:AJ400630 hypothetical protein [Streptococcus pneumoniae bacteriophage MM1]
 Identities = 254/321 (79%), Positives = 286/321 (88%), Gaps = 1/321 (0%)

Query: 1   MRKPKITVIGGGTGIPVILKSLRLEDVEITAVVTVADDGGSSGELRSVMQ-LTPPGDLRN 59
            MRKPKITVIGGGTGIPVILKSLR +DVEI A+VTVADDGGSSGELR  MQ LTPPGDLRN
Sbjct: 1   MRKPKITVIGGGTGIPVILKSLREKDVEIAAIVTVADDGGSSGELRKNMQQLTPPGDLRN 60

Query: 60  VLVALSDMPKFYEQIFQYRFAEGDGDFAGHPLGNLIIAGVAEMQGSTYNAMQSLTQFFHT 119
            VLVA+SDMPKFYE++FQYRF+E  G FAGHPLGNLIIAG++EMQGSTYNAMQ L++FFHT
Sbjct: 61  VLVAMSDMPKFYEKVFQYRFSEDAGAFAGHPLGNLIIAGLSEMQGSTYNAMQLLSKFFHT 120

Query: 120 TGKIYPSSEHPLTLHAVFKDGHEVVGESQIADYKGMIDHVYVTNTYNEETPTASRKVVDA 179
            TGKIYPSS+HPLTLHAVF+DG EV GES I D++G+ID+VYVTN  N++TP ASR+VV
Sbjct: 121 TGKIYPSSDHPLTLHAVFQDGTEVAGESHIVDHRGIIDNVYVTNALNDDTPLASRRVVQT 180

Query: 180 ILESDMIVLGPGSLFTSILPNLVIPEIKQALLETRAEVAYVCNIMTQRGETEHFTDADHV 239
            ILESDMIVLGPGSLFTSILPN+VI EI +ALLET+AE+AYVCNIMTQRGETEHFTD+DHV
Sbjct: 181 ILESDMIVLGPGSLFTSILPNIVIKEIGRALLETKAEIAYVCNIMTQRGETEHFTDSDHV 240

Query: 240 EVLKRHLGQDAIDTVLVNIEKVPESYMENNHFDEYLVQVEHDFSGLRKHARRVISSNFLK 299
            EVL RHLG+  IDTVLVNIEKVP+ YM +N FDEYLVQVEHDF GL K   RVISSNFL+
Sbjct: 241 EVLHRHLGRPFIDTVLVNIEKVPQEYMNSNRFDEYLVQVEHDFVGLCKQVSRVISSNFLR 300

Query: 300 LEKGGAFHHGDFVVEELMNLV 320
            LE GGAFH GD +V+ELM ++
Sbjct: 301 LENGGAFHDGDLIVDELMRII 321
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5535> which encodes the amino acid sequence <SEQ ID 5536>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 251/320 (78%), Positives = 284/320 (88%)

Query: 1   MRKPKITVIGGGTGIPVILKSLRLEDVEITAVVTVADDGGSSGELRSVMQLTPPGDLRNV 60
            M+ PK+TVIGGGTGI +ILKSLR E V+ITAVVTVADDGGSSGELR+ MQL PPGDLRNV
Sbjct: 1   MKNPKMTVIGGGTGISIILKSLRNEAVDITAVVTVADDGGSSGELRNAMQLAPPGDLRNV 60

Query: 61  LVALSDMPKFYEQIFQYRFAEGDGDFAGHPLGNLIIAGVAEMQGSTYNAMQSLTQFFHTT 120
```

```
           L+A+SDMPKFYE++FQYRF E DG  AGHPLGNLIIAG++EMQGSTYNA+Q LT+FFH T
Sbjct: 61  LLAMSDMPKFYERVFQYRFNESDGALAGHPLGNLIIAGISEMQGSTYNAIQILTKFFHIT 120

Query: 121 GKIYPSSEHPLTLHAVFKDGHEVVGESQIADYKGMIDHVYVTNTYNEETPTASRKVVDAI 180
           GKIYPSSE  LTLHAVFKDGHEV GES IA Y GMIDHVYVTNTYN++ P ASRKVV+AI
Sbjct: 121 GKIYPSSEQALTLHAVFKDGHEVAGESSIAKYPGMIDHVYVTNTYNDQKPQASRKVVEAI 180

Query: 181 LESDMIVLGPGSLFTSILPNLVIPEIKQALLETRAEVAYVCNIMTQRGETEHFTDADHVE 240
           LESDMIVLGPGSLFTSILPNLVIPEIK+AL +T+AEV Y+CNIMTQ GETE F+DADHV
Sbjct: 181 LESDMIVLGPGSLFTSILPNLVIPEIKEALRQTKAEVVYICNIMTQYGETEQFSDADHVA 240

Query: 241 VLKRHLGQDAIDTVLVNIEKVPESYMENNHFDEYLVQVEHDFSGLRKHARRVISSNFLKL 300
           VL +HLG+D IDTVLVN+ KVP++YM +N FDEYLVQV+HDF+GL + A+RVISS FL+L
Sbjct: 241 VLNQHLGRDLIDTVLVNVAKVPQAYMNSNKFDEYLVQVDHDFAGLCRAAKRVISSYFLRL 300

Query: 301 EKGGAFHHGDFVVEELMNLV 320
           E GGAFH G+ VVEELMNLV
Sbjct: 301 ENGGAFHDGNLVVEELMNLV 320
```

SEQ ID 5534 (GBS269) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 12; MW 35kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 54 (lane 5; MW 60.5kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1781

A DNA sequence (GBSx1888) was identified in *S.agalactiae* <SEQ ID 5537> which encodes the amino acid sequence <SEQ ID 5538>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2479(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB96619 GB:AJ400630 hypothetical protein [Streptococcus pneumoniae bacteriophage MM1]
 Identities = 209/303 (68%), Positives = 260/303 (84%)

Query: 1   MSFTVKVKEELLGHKSENKMELSAIIKMSGSLGLANHGLNLSITTENAKIARHIYSMLEE 60
           MSFTV VKEE+LG    ++ ELSAIIKMSGS+GL+  GL LS+ TENAK+ARH+Y
Sbjct: 1   MSFTVAVKEEILGQHHLSRHELSAIIKMSGSIGLSTSGLTLSVVTENAKLARHLYESFLH 60

Query: 61  HYHLQPEIKYHQKTNLRKNRVYTVFIEEKVDVILADLKLADAFFGIETGIEHSILDNDEN 120
            Y ++ EI++HQ++NLRKNRVYTVF +EKV  +L+DL LAD+FFG+ETGI+ +IL ++E
Sbjct: 61  FYEIKSEIRHHQRSNLRKNRVYTVFTDEKVQDLLSDLHLADSFFGLETGIDEAILSDEEA 120

Query: 121 GRAYLRGAFLSTGTVREPDSGKYQLEIFSVYLDHAQDLANLMKKFMLDAKVIEHKHGAVT 180
           GRAYL GAFL+ G++R+P+SGKYQLEI SVYLDHAQ +A+L+++F+LDAKV+E K GAVT
Sbjct: 121 GRAYLCGAFLANGSIRDPESGKYQLEISSVYLDHAQGIASLLQQFLLDAKVLERKKGAVT 180

Query: 181 YLQKAEDIMDFLIVIDAMEARDAFEEIKMIRETRNDINRANNVETANIARTITASMKTIN 240
           YLQ+AEDIMDFLIVI AM+ARD FE +K++RETRND+NRANN ETANIART++ASMKTIN
Sbjct: 181 YLQRAEDIMDFLIVIGAMQARDDFERVKILRETRNDLNRANNAETANIARTVSASMKTIN 240

Query: 241 NIIKIMDTIGFDALPSDLRQVAQVRVAHPDYSIQQIADSLETPLSKSGVNHRLRKINKIA 300
           NI KI D +G + LP DL++VAQ+R+ HPDYSIQQ+ADSL TPL+KSGVNHRLRKINKIA
Sbjct: 241 NISKIKDIMGLENLPVDLQEVAQLRIQHPDYSIQQLADSLSTPLTKSGVNHRLRKINKIA 300
```

```
Query: 301 DEL 303
           DEL
Sbjct: 301 DEL 303
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5539> which encodes the amino acid sequence <SEQ ID 5540>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1698(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 222/303 (73%), Positives = 269/303 (88%)

Query: 1   MSFTVKVKEELLGHKSENKMELSAIIKMSGSLGLANHGLNLSITTENAKIARHIYSMLEE 60
           MSFT KVKEEL+    + +  EL+AIIK+SGSLGLA+  L+LSITTENAKIAR+IYS++E+
Sbjct: 1   MSFTTKVKEELIHLSTGDNNELAAIIKLSGSLGLAHQSLHLSITTENAKIARYIYSLIED 60

Query: 61  HYHLQPEIKYHQKTNLRKNRVYTVFIEEKVDVILADLKLADAFFGIETGIEHSILDNDEN 120
            Y + PEI+YHQKTNLRKNRVYTV++E+ V+ ILADLKLAD+FFG+ETGIE  +L +D
Sbjct: 61  AYVIVPEIRYHQKTNLRKNRVYTVYVEQGVETILADLKLADSFFGLETGIEPQVLSDDNA 120

Query: 121 GRAYLRGAFLSTGTVREPDSGKYQLEIFSVYLDHAQDLANLMKKFMLDAKVIEHKHGAVT 180
           GR+YL+GAFL+ G++R+P+SGKYQLEI+SVYLDHAQDLA LM+KFMLDAK IEHK GAVT
Sbjct: 121 GRSYLKGAFLAAGSIRDPESGKYQLEIYSVYLDHAQDLAQLMQKFMLDAKTIEHKSGAVT 180

Query: 181 YLQKAEDIMDFLIVIDAMEARDAFEEIKMIRETRNDINRANNVETANIARTITASMKTIN 240
           YLQKAEDIMDFLI+I AM  ++ FE IK++RE RNDINRANN ETANIA+TI+ASMKTIN
Sbjct: 181 YLQKAEDIMDFLIIIGAMSCKEDFEAIKLLREARNDINRANNAETANIAKTISASMKTIN 240

Query: 241 NIIKIMDTIGFDALPSDLRQVAQVRVAHPDYSIQQIADSLETPLSKSGVNHRLRKINKIA 300
           NIIKIMDTIG ++LP +L+QVAQ+RV HPDYSIQQ+AD+LE P++KSGVNHRLRKINKIA
Sbjct: 241 NIIKIMDTIGLESLPIELQQVAQLRVKHPDYSIQQVADALEFPITKSGVNHRLRKINKIA 300

Query: 301 DEL 303
           D+L
Sbjct: 301 DDL 303
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1782

A DNA sequence (GBSx1889) was identified in *S.agalactiae* <SEQ ID 5541> which encodes the amino acid sequence <SEQ ID 5542>. This protein is predicted to be dipeptidase. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3544(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA86210 GB:Z38063 dipeptidase [Lactobacillus helveticus]
 Identities = 218/473 (46%), Positives = 310/473 (65%), Gaps = 14/473 (2%)
```

```
Query: 3   CTTILVGKKASYDGSTMIARTEDSVNGDFTPKKLKVMTSKDQPRHYKSVLSNFEVD---L 59
           CTTILVGKKAS DGSTMIAR+ED       P+  KV+  +DQP+HY SV+S  ++D   L
Sbjct: 6   CTTILVGKKASIDGSTMIARSEDG-GRVIIPEGFKVVNPEDQPKHYTSVISKQKIDDEDL 64

Query: 60  PDNPLPYTSVPDALGKDGIWGEAGINSKNVAMSATETITTNSRVLGADPLVSD---GIGE 116
            + PL YTS PD  GK+GIWG AGIN+ NVAM+ATETITTNSR+ G DP++     G+GE
Sbjct: 65  AETPLRYTSAPDVSGKNGIWGAAGINADNVAMTATETITTNSRIQGVDPILDPSEGGLGE 124

Query: 117 EDILTLVLPYIQSAREGVERLGAILEKYGTYESNGIAFSDTEEIWWLETIGGHHWIARRV 176
           ED +TL LPY+ SA +GV+R+G ++EKYGTYE NG+AFSD + IW+LETIGGHHWIARR+
Sbjct: 125 EDFVTLTLPYLHSAFDGVKRVGYLVEKYGTYEMNGMAFSDKDNIWYLETIGGHHWIARRI 184

Query: 177 PDDVYVTNPNQLGIDHFEFNNCDDYMCSSDLKEFIEQYHLDLTYSNEHFNPRYAFGSQRD 236
           PDD YV  PN+L ID F+F++ +++  +SDLK+ I++YHL+     E +N R+ FGS
Sbjct: 185 PDDAYVIAPNRLNIDTFDFDDSENFAAASDLKDLIDEYHLN--PDREGYNMRHIFGSSTI 242

Query: 237 KDRHYNTPRSWAMQRFLNPEIEQDPRSLFIPWCQKPYRKITVEDIKYVLSDHYQDSVYDP 296
           KD HYN PR+W +  + +P+    P     P+  +  R I++EDIK+  S HYQD+ YD
Sbjct: 243 KDAHYNNPRAWYIHNYFDPDFGGTPADQDQPFICRANRLISIEDIKWAESSHYQDTPYDA 302

Query: 297 YGPEGDAVSRRAFRSVGINRTSQTSILQLRPNKSLETTGVQWLSYGSMPFATMVPLFTQV 356
           YG +G    ++ FR +GINR  +T ILQ+R +   E  GVQWL++G   F +M+P +T V
Sbjct: 303 YGDQGTPEQKKTFRPIGINRNFETHILQIRNDVPAEIAGVQWLAFGPNTFNSMLPFYTNV 362

Query: 357 ETVPNYFSNTTKDASTDNFYWTNRLIAALADPHFYQHEADIESYIERTMAQGHAHINGVD 416
            T P  +  T K  + +  +W N+L A L D ++  +     +++ ++++AQ H   +  D
Sbjct: 363 TTTPEAWQTTPK-FNLNKIFWLNKLTAQLGDTNYRVYGELEDAFEQKSLAQCHKIQHETD 421

Query: 417 REVAENKEIDFQQK----NQEMSDYIQKESQELLNRILFDASNLMTNRFSMGD 465
           +EV      + Q K    NQ+MSD +   + ELL +++ +   LMT ++ + D
Sbjct: 422 KEVKNLSGKELQDKLIAANQKMSDTVYNNTVELLGQMVDEGHGLMTLKYDLLD 474
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5543> which encodes the amino acid sequence <SEQ ID 5544>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.0514(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 345/464 (74%), Positives = 407/464 (87%)

Query: 2   ACTTILVGKKASYDGSTMIARTEDSVNGDFTPKKLKVMTSKDQPRHYKSVLSNFEVDLPD 61
           +CTTILVGKKASYDGSTM+ARTEDS NGDFTPKK+ V+  +DQPRHY+SV S+FE+DLPD
Sbjct: 9   SCTTILVGKKASYDGSTMVARTEDSQNGDFTPKKMIVVKPEDQPRHYRSVQSSFEMDLPD 68

Query: 62  NPLPYTSVPDALGKDGIWGEAGINSKNVAMSATETITTNSRVLGADPLVSDGIGEEDILT 121
           NP+ YTSVPDALGKDGIW EAG+N  NVAMSATETITTNSRVLGADPLV+ GIGEED++T
Sbjct: 69  NPMTYTSVPDALGKDGIWAEAGVNEANVAMSATETITTNSRVLGADPLVASGIGEEDMVT 128

Query: 122 LVLPYIQSAREGVERLGAILEKYGTYESNGIAFSDTEEIWWLETIGGHHWIARRVPDDVY 181
           LVLPYI+SAREGV RLGAILE YGTYESNG+AFSD  +IWWLETIGGHHWIARRVPDD Y
Sbjct: 129 LVLPYIRSAREGVLRLGAILEDYGTYESNGVAFSDEHDIWWLETIGGHHWIARRVPDDAY 188

Query: 182 VTNPNQLGIDHFEFNNCDDYMCSSDLKEFIEQYHLDLTYSNEHFNPRYAFGSQRDKDRHY 241
           VTNPNQ GIDHFEFNN +DY+CS+DLK+FI+ YHLDLTYS+EHFNPRYAFGSQRDKDR Y
Sbjct: 189 VTNPNQFGIDHFEFNNPEDYLCSADLKDFIDTYHLDLTYSHEHFNPRYAFGSQRDKDRQY 248

Query: 242 NTPRSWAMQRFLNPEIEQDPRSLFIPWCQKPYRKITVEDIKYVLSDHYQDSVYDPYGPEG 301
           NTPR+W MQ+FLNPEI QDPRS  + WCQKPYRKITVED+KYVLS HYQD+ YDPYG EG
Sbjct: 249 NTPRAWIMQKFLNPEIVQDPRSFALAWCQKPYRKITVEDVKYVLSSHYQDTGYDPYGSEG 308
```

```
Query: 302 DAVSRRAFRSVGINRTSQTSILQLRPNKSLETTGVQWLSYGSMPFATMVPLFTQVETVPN 361
             VS++ FR +GINRTSQT+IL +RPNK  E   +QW++YGSMPF TMVP FTQV+T+P+
Sbjct: 309 TPVSKKVFRPIGINRTSQTAILHIRPNKPQEIAAIQWMAYGSMPFNTMVPFFTQVKTIPD 368

Query: 362 YFSNTTKDASTDNFYWTNRLIAALADPHFYQHEADIESYIERTMAQGHAHINGVDREVAE 421
           YF+NT ++  TDNFYWTNRLIAALADPH+  HE D+++Y+E TMA+GHA ++ V+ ++
Sbjct: 369 YFANTYENVFTDNFYWTNRLIAALADPHYNHHETDLDNYLEETMAKGHAMLHAVEVQLLA 428

Query: 422 NKEIDFQQKNQEMSDYIQKESQELLNRILFDASNLMTNRFSMGD 465
            + +D +++NQ+MSDY+Q E+Q LLN+ILFDASNLMTNRFS+ D
Sbjct: 429 GETVDLEEENQKMSDYVQGETQTLLNKILFDASNLMTNRFSLSD 472
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1783

A DNA sequence (GBSx1890) was identified in *S.agalactiae* <SEQ ID 5545> which encodes the amino acid sequence <SEQ ID 5546>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA96185 GB:Z71552 AdcA protein [Streptococcus pneumoniae]
 Identities = 257/429 (59%), Positives = 312/429 (71%), Gaps = 7/429 (1%)

Query: 1   MRKKFLLLMSFVAMFAAWQLVQVKQVWADSKLKVVTTFYPVYEFTKNVVGDKADVSMLIK 60
           M+K  LLL S  A+F    +    Q  AD KL +VTTFYPVYEFTK V GD A+V +LI
Sbjct: 1   MKKISLLLASLCALFL---VACSNQKQADGKLNIVTTFYPVYEFTKQVAGDTANVELLIG 57

Query: 61  AGTEPHDFEPSTKNIAAIQDSNAFVYMDDNMETWAPKVAKSVKSKKVTTIKGTGDMLLTK 120
           AGTEPH++EPS K +A IQD++ FVY ++NMETW PK+  ++  KKV TIK TGDMLL
Sbjct: 58  AGTEPHEYEPSAKAVAKIQDADTFVYENENMETWVPKLLDTLDKKKVKTIKATGDMLLLP 117

Query: 121 GVEEEGEEHEGHGHEGHHHELDPHVWLSPERAISVVENIRNKFVKAYPKDAASFNKNADA 180
           G EEE  +H+ HG EGHHHE DPHVWLSP RAI +VE+IR+     YP    +F KNA A
Sbjct: 118 GGEEEEGDHD-HGEEGHHHEFDPHVWLSPVRAIKLVEHIRDTLSADYPDKKETFEKNAAA 176

Query: 181 YIAKLKELDKEYKNGLSNAKQKSFVTQHAAFGYMALDYGLNQVPIAGLTPDAEPSSKRLG 240
           YI KL+ LDK Y  GLS AK+KSFVTQHAAF Y+ALDYGL QV I+GL+PDAEPS+ RL
Sbjct: 177 YIEKLQSLDKAYAEGLSQAKEKSFVTQHAAFNYLALDYGLKQVAISGLSPDAEPSAARLA 236

Query: 241 ELAKYIKKYNINYIYFEENASNKVAKTLADEVGVKTAVLSPLEGLSKKEMAAGEDYFSVM 300
           EL +Y+KK  I YIYFEENAS  +A TL+ E GVKT VL+PLE L++++  AGE+Y SVM
Sbjct: 237 ELTEYVKKNKIAYIYFEENASQALANTLSKEAGVKTDVLNPLESLTEEDTKAGENYISVM 296

Query: 301 RRNLKVLKKTTDVAGKEVAPEE-DKTKTVETGYFKTKDVKDRKLTDYSGNWQSVYPLLQD 359
            +NLK LK+TTD  G  + PE+ + TKTV+ GYF+    VKDR L+DY+GNWQSVYP L+D
Sbjct: 297 EKNLKALKQTTDQEGPAIEPEKAEDTKTVQNGYFEDAAVKDRTLSDYAGNWQSVYPFLED 356

Query: 360 GTLDPVWDYKAKSKKDMTAAEYKKYYTAGYKTDVESIKIDGKKHQMTFVRNGKSQTFTYK 419
           GT D V+DYKAK    MT AEYK YYT GY+TDV  I I    + M FV+ G+S+ +TYK
Sbjct: 357 GTFDQVFDYKAKLTGKMTQAEYKAYYTKGYQTDVTKINI--TDNTMEFVQGGQSKKYTYK 414

Query: 420 YAGYKILTY 428
           Y G KILTY
Sbjct: 415 YVGKKILTY 423
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5547> which encodes the amino acid sequence <SEQ ID 5548>. Analysis of this protein sequence reveals the following:

```
        Possible site: 17

   >>> Seems to have a cleavable N-term signal seq.

   ----- Final Results -----
                  bacterial outside --- Certainty=0.3000(Affirmative) < succ>
                 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA96185 GB:Z71552 AdcA protein [Streptococcus pneumoniae]
 Identities = 259/438 (59%), Positives = 326/438 (74%), Gaps = 16/438 (3%)

Query: 1   MKKKILLMMSLISVFFAWQLTQAKQVLAEGKVKVVTTFYPVYEFTKGVIGNDGDVFMLMK 60
           MKK  LL+ SL ++F    +  + Q  A+GK+ +VTTFYPVYEFTK V G+  +V +L+
Sbjct: 1   MKKISLLLASLCALFL---VACSNQKQADGKLNIVTTFYPVYEFTKQVAGDTANVELLIG 57

Query: 61  AGTEPHDFEPSTKDIKKIQDADAFVYMDDNMETWVSDVKKSLTSKKVTIVKGTGNMLLVA 120
           AGTEPH++EPS K + KIQDAD FVY ++NMETWV  +  +L  KKV  +K TG+MLL+
Sbjct: 58  AGTEPHEYEPSAKAVAKIQDADTFVYENENMETWVPKLLDTLDKKKVKTIKATGDMLLLP 117

Query: 121 GAGHDHPHEDADKKHEHNKHSEEGHNHAFDPHVWLSPYRSITVVENIRDSLSKAYPEKAE 180
           G       E+ +  H+H    EEGH+H FDPHVWLSP R+I +VE+IRD+LS  YP+K E
Sbjct: 118 GG------EEEEGDHDHG---EEGHHHEFDPHVWLSPVRAIKLVEHIRDTLSADYPDKKE 168

Query: 181 NFKANAATYIEKLKELDKDYTAALSDAKQKSFVTQHAAFGYMALDYGLNQISINGVTPDA 240
            F+ NAA YIEKL+ LDK Y   LS AK+KSFVTQHAAF Y+ALDYGL Q++I+G++PDA
Sbjct: 169 TFEKNAAAYIEKLQSLDKAYAEGLSQAKEKSFVTQHAAFNYLALDYGLKQVAISGLSPDA 228

Query: 241 EPSAKRIATLSKYVKKYGIKYIYFEENASSKVAKTLAKEAGVKAAVLSPLEGLTEKEMKA 300
           EPSA R+A L++YVKK  I YIYFEENAS  +A TL+KEAGVK  VL+PLE LTE++ KA
Sbjct: 229 EPSAARLAELTEYVKKNKIAYIYFEENASQALANTLSKEAGVKTDVLNPLESLTEEDTKA 288

Query: 301 GQDYFTVMRKNLETLRLTTDVAGKEILPEK-DTTKTVYNGYFKDKEVKDRQLSDWSGSWQ 359
           G++Y +VM KNL+ L+ TTD  G  I PEK + TKTV NGYF+D  VKDR LSD++G+WQ
Sbjct: 289 GENYISVMEKNLKALKQTTDQEGPAIEPEKAEDTKTVQNGYFEDAAVKDRTLSDYAGNWQ 348

Query: 360 SVYPYLQDGTLDQVWDYKAKKSKGKMTAAEYKDYYTTGYKTDVEQIKINGKKKTMTFVRN 419
           SVYP+L+DGT DQV+DYKAK + GKMT AEYK YYT GY+TDV   KIN    TM FV+
Sbjct: 349 SVYPFLEDGTFDQVFDYKAKLT-GKMTQAEYKAYYTKGYQTDV--TKINITDNTMEFVQG 405

Query: 420 GEKKTFTYTYAGKEILTY 437
           G+ K +TY Y GK+ILTY
Sbjct: 406 GQSKKYTYKYVGKKILTY 423
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 353/515 (68%), Positives = 422/515 (81%), Gaps = 9/515 (1%)

Query: 1   MRKKFLLLMSFVAMFAAWQLVQVKQVWADSKLKVVTTFYPVYEFTKNVVGDKADVSMLIK 60
           M+KK LL+MS +++F AWQL Q KQV A+ K+KVVTTFYPVYEFTK V+G+  DV ML+K
Sbjct: 1   MKKKILLMMSLISVFFAWQLTQAKQVLAEGKVKVVTTFYPVYEFTKGVIGNDGDVFMLMK 60

Query: 61  AGTEPHDFEPSTKNIAAIQDSNAFVYMDDNMETWAPKVAKSVKSKKVTTIKGTGDMLLTK 120
           AGTEPHDFEPSTK+I  IQD++AFVYMDDNMETW   V KS+ SKKVT +KGTG+MLL
Sbjct: 61  AGTEPHDFEPSTKDIKKIQDADAFVYMDDNMETWVSDVKKSLTSKKVTIVKGTGNMLLVA 120

Query: 121 GV--------EEEGEEHEGHGHEGHHHELDPHVWLSPERAISVVENIRNKFVKAYPKDAA 172
           G          ++  EH  H  EGH+H  DPHVWLSP R+I+VVENIR+   KAYP+ A
Sbjct: 121 GAGHDHPHEDADKKHEHNKHSEEGHNHAFDPHVWLSPYRSITVVENIRDSLSKAYPEKAE 180

Query: 173 SFNKNADAYIAKLKELDKEYKNGLSNAKQKSFVTQHAAFGYMALDYGLNQVPIAGLTPDA 232
           +F  NA  YI KLKELDK+Y   LS+AKQKSFVTQHAAFGYMALDYGLNQ+ I G+TPDA
```

```
Sbjct: 181 NFKANAATYIEKLKELDKDYTAALSDAKQKSFVTQHAAFGYMALDYGLNQISINGVTPDA 240

Query: 233 EPSSKRLGELAKYIKKYNINYIYFEENASNKVAKTLADEVGVKTAVLSPLEGLSKKEMAA 292
           EPS+KR+  L+KY+KKY I YIYFEENAS+KVAKTLA E GVK AVLSPLEGL++KEM A
Sbjct: 241 EPSAKRIATLSKYVKKYGIKYIYFEENASSKVAKTLAKEAGVKAAVLSPLEGLTEKEMKA 300

Query: 293 GEDYFSVMRRNLKVLKKTTDVAGKEVAPEEDKTKTVETGYFKTKDVKDRKLTDYSGNWQS 352
           G+DYF+VMR+NL+ L+ TTDVAGKE+ PE+D TKTV  GYFK K+VKDR+L+D+SG+WQS
Sbjct: 301 GQDYFTVMRKNLETLRLTTDVAGKEILPEKDTTKTVYNGYFKDKEVKDRQLSDWSGSWQS 360

Query: 353 VYPLLQDGTLDPVWDYKA-KSKKDMTAAEYKKYYTAGYKTDVESIKIDGKKHQMTFVRNG 411
           VYP LQDGTLD VWDYKA KSK  MTAAEYK YYT GYKTDVE IKI+GKK  MTFVRNG
Sbjct: 361 VYPYLQDGTLDQVWDYKAKKSKGKMTAAEYKDYYTTGYKTDVEQIKINGKKKTMTFVRNG 420

Query: 412 KSQTFTYKYAGYKILTYKKGNRGVRYLFEAKEKDAGQFKYIQFSDHGIKPNKAEHFHIFW 471
           + +TFTY YAG +ILTY KGNRGVR++FEAKE DAG+FKY+QFSDH I P KA+HFH++W
Sbjct: 421 EKKTFTYTYAGKEILTYPKGNRGVRFMFEAKEADAGEFKYVQFSDHAIAPEKAKHFHLYW 480

Query: 472 GSESQEKLFEEMENWPTYFPAKMSGREVAQDLMSH 506
           G +SQEKL +E+E+WPTY+ + +SGRE+AQ++ +H
Sbjct: 481 GGDSQEKLHKELEHWPTYYGSDLSGREIAQEINAH 515
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8899> and protein <SEQ ID 8900> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 3
SRCFLG: 0
McG: Length  of  UR:   19
     Peak Value of UR:   2.79
     Net Charge of CR: 3
McG: Discrim Score:      9.08
GvH: Signal Score (-7.5): 2.59
     Possible site: 15
>>> Seems to have a cleavable N-term signal seq.
Amino Acid Composition: calculated from 16
ALOM program   count: 0 value:   7.69 threshold:  0.0
   PERIPHERAL  Likelihood =  7.69     264
 modified ALOM score:  -2.04

*** Reasoning Step: 3

Rule gpo1

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
3758895|emb|CAA96185.1||Z71552 AdcA protein {Streptococcus pneumoniae} >PIR|T46756|T46756
Zn-binding lipoprotein
           adcA [imported] - Streptococcus pneumoniae (fragment)

 Score =  508 bits (1294), Expect = e-143
 Identities = 257/429 (59%), Positives = 312/429 (71%), Gaps = 7/429 (1%)

Query: 1    MRKKFLLLMSFVAMFAAWQLVQVKQVWADSKLKVVTTFYPVYEFTKNVVGDKADVSMLIK 60
            M+K  LLL S  A+F    +    Q  AD KL +VTTFYPVYEFTK V GD A+V +LI
Sbjct: 1    MKKISLLLASLCALFL---VACSNQKQADGKLNIVTTFYPVYEFTKQVAGDTANVELLIG 57

Query: 61   AGTEPHDFEPSTKNIAAIQDSNAFVYMDDNMETWAPKVAKSVKSKKVTTIKGTGDMLLTK 120
            AGTEPH++EPS K +A IQD++ FVY ++NMETW PK+  ++  KKV TIK TGDMLL
```

```
Sbjct: 58  AGTEPHEYEPSAKAVAKIQDADTFVYENENMETWVPKLLDTLDKKKVKTIKATGDMLLLP 117

Query: 121 GVEEEGEEHEGHGHEGHHHELDPHVWLSPERAISVVENIRNKFVKAYPKDAASFNKNADA 180
           G EEE  +H+ HG EGHHHE DPHVWLSP RAI +VE+IR+     YP    +F KNA A
Sbjct: 118 GGEEEEGDHD-HGEEGHHHEFDPHVWLSPVRAIKLVEHIRDTLSADYPDKKETFEKNAAA 176

Query: 181 YIAKLKELDKEYKNGLSNAKQKSFVTQHAAFGYMALDYGLNQVPIAGLTPDAEPSSKRLG 240
           YI KL+ LDK Y  GLS AK+KSFVTQHAAF Y+ALDYGL QV I+GL+PDAEPS+ RL
Sbjct: 177 YIEKLQSLDKAYAEGLSQAKEKSFVTQHAAFNYLALDYGLKQVAISGLSPDAEPSAARLA 236

Query: 241 ELAKYIKKYNINYIYFEENASNKVAKTLADEVGVKTAVLSPLEGLSKKEMAAGEDYFSVM 300
           EL +Y+KK  I YIYFEENAS  +A TL+ E GVKT VL+PLE L++++  AGE+Y SVM
Sbjct: 237 ELTEYVKKNKIAYIYFEENASQALANTLSKEAGVKTDVLNPLESLTEEDTKAGENYISVM 296

Query: 301 RRNLKVLKKTTDVAGKEVAPEE-DKTKTVETGYFKTKDVKDRKLTDYSGNWQSVYPLLQD 359
            +NLK LK+TTD  G  + PE+ + TKTV+ GYF+    VKDR L+DY+GNWQSVYP L+D
Sbjct: 297 EKNLKALKQTTDQEGPAIEPEKAEDTKTVQNGYFEDAAVKDRTLSDYAGNWQSVYPFLED 356

Query: 360 GTLDPVWDYKAKSKKDMTAAEYKKYYTAGYKTDVESIKIDGKXHQMTFVRNGKSQTFTYK 419
           GT D V+DYKAK    MT AEYK YYT GY+TDV  I I    + M FV+ G+S+ +TYK
Sbjct: 357 GTFDQVFDYKAKLTGKMTQAEYKAYYTKGYQTDVTKINI--TDNTMEFVQGGQSKKYTYK 414

Query: 420 YAGYKILTY 428
           Y G KILTY
Sbjct: 415 YVGKKILTY 423
```

SEQ ID 8900 (GBS325) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 65 (lane 3; MW 58kDa).

The GBS325-His fusion product was purified (Figure 210, lane 7) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 257A) and FACS (Figure 257B). These tests confirm that the protein is immunoaccessible on GBS bacteria.

Example 1784

A DNA sequence (GBSx1891) was identified in *S.agalactiae* <SEQ ID 5549> which encodes the amino acid sequence <SEQ ID 5550>. This protein is predicted to be ribosomal protein L31 (rl31). Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1948(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9637> which encodes amino acid sequence <SEQ ID 9638> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF80389 GB:AF160251 ribosomal protein L31 [Listeria innocua]
 Identities = 61/81 (75%), Positives = 71/81 (87%), Gaps = 1/81 (1%)

Query: 9   MKKDIHPDYRPVVFLDTTTGYKFLSGSTKSTKETVEFE-GETYPLIRVEISSDSHPFYTG 67
           MK  IHP+YRPVVF+DT+T +KFLSGSTKS+ ET+++E G  YPL+RVEISSDSHPFYTG
Sbjct: 1   MKTGIHPEYRPVVFVDTSTDFKFLSGSTKSSSETIKWEDGNEYPLLRVEISSDSHPFYTG 60

Query: 68 RQKFTQADGRVDRFNKKYGLK 88
           +QK   ADGRVDRFNKKYGLK
Sbjct: 61 KQKHATADGRVDRFNKKYGLK 81
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5551> which encodes the amino acid sequence <SEQ ID 5552>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1910(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 81/86 (94%), Positives = 86/86 (99%)

Query: 9  MKKDIHPDYRPVVFLDTTTGYKFLSGSTKSTKETVEFEGETYPLIRVEISSDSHPFYTGR 68
           M+KDIHPDYRPVVFLDTTTGY+FLSGSTK++KETVEFEGETYPLIRVEISSDSHPFYTGR
Sbjct: 1  MRKDIHPDYRPVVFLDTTTGYQFLSGSTKASKETVEFEGETYPLIRVEISSDSHPFYTGR 60

Query: 69 QKFTQADGRVDRFNKKYGLKDANAAQ 94
           QKFTQADGRVDRFNKKYGLKDANAA+
Sbjct: 61 QKFTQADGRVDRFNKKYGLKDANAAK 86
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1785

A DNA sequence (GBSx1892) was identified in *S.agalactiae* <SEQ ID 5553> which encodes the amino acid sequence <SEQ ID 5554>. This protein is predicted to be aspartate aminotransferase (aspC). Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1740(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9421> which encodes amino acid sequence <SEQ ID 9422> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC21948 GB:U32714 aminotransferase [Haemophilus influenzae Rd]
 Identities = 200/323 (61%), Positives = 264/323 (80%), Gaps = 1/323 (0%)

Query: 1   MQYYQLQNI-HVDMDDIYIVNGVSEGISMSMQALLDNDDEVLVPMPDYPLWTACVSLAGG 59
            +QYYQ + I    ++D+YI NGVSE I+M+MQALL++ DEVLVPMPDYPLWTA V+L+GG
Sbjct: 82  VQYYQSKGILGATVNDVYIGNGVSELITMAMQALLNDGDEVLVPMPDYPLWTAAVTLSGG 141

Query: 60  NAVHYICDEEANWYPDIDDIKSKITSKTKAIVLINPNNPTGAVYPREILQEIVDIARQND 119
             AVHY+CDE+ANW+P IDDIK+K+ +KTKAIV+INPNNPTGAVY +E+LQEIV+IARQN+
Sbjct: 142 KAVHYLCDEDANWFPTIDDIKAKVNAKTKAIVIINPNNPTGAVYSKELLQEIVEIARQNN 201

Query: 120 LIIFSDEVYDRLVMDGMEHIPIASIAEDIFTVTLSGLSKSHRICGFRVGWMVLSGPRQHV 179
            LIIF+DE+YD+++ DG  H  IA++A D+ TVTL+GLSK++R+ GFR GWM+L+GP+ +
Sbjct: 202 LIIFADEIYDKILYDGAVHHHIAALAPDLLTVTLNGLSKAYRVAGFRQGWMILNGPKHNA 261

Query: 180 KGYIEGLNMLANMRLCSNVLAQQVIQTSLGGQQSIDSMLLPGGRIYEQRNYIHKAINEIP 239
            KGYIEGL+MLA+MRLC+NV  Q  IQT+LGG QSI+  +LPGGR+ EQRN  +  I +IP
Sbjct: 262 KGYIEGLDMLASMRLCANVPMQHAIQTALGGYQSINEFILPGGRLLEQRNKAYDLITQIP 321
```

```
Query: 240 GLSAVKPNAGLYLFPKIDTDMYRIDNDEEFVLNFLKQEKVLLTHGRGFNMNTADHFRIVY 299
           G++ VKP   +Y+FPKID   + I +DE+ VL+ L+QEKVLL HG+GFN ++ DHFRIV
Sbjct: 322 GITCVKPMGAMYMFPKIDVKKFNIHSDEKMVLDLLRQEKVLLVHGKGFNWHSPDHFRIVT 381

Query: 300 LPRVDELTELQEKMARFLSQYKR 322
           LP V++L E   K+ARFLS Y++
Sbjct: 382 LPYVNQLEEAITKLARFLSDYRQ 404
```

There is also homology to SEQ ID 3662.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1786

A DNA sequence (GBSx1893) was identified in *S.agalactiae* <SEQ ID 5555> which encodes the amino acid sequence <SEQ ID 5556>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.02    Transmembrane  164 - 180 ( 163 - 181)

----- Final Results -----
               bacterial membrane --- Certainty=0.1808(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10099> which encodes amino acid sequence <SEQ ID 10100> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06181 GB:AP001515 transcriptional pleiotropic repressor
          [Bacillus halodurans]
 Identities = 129/257 (50%), Positives = 181/257 (70%), Gaps = 3/257 (1%)

Query: 23  NLLEKTRKITSILQRSVDSLDAELPYNTMAAQLADIIDCNACIINGGGNLLGYAMKYKTN 82
           +LL + RKI  +LQ+S       + +  MA  L D+I  N  +++  G LLG+A+K +
Sbjct: 2   SLLSRMRKINDMLQKSGVQ---HVNFREMAETLRDVISANIFVVSRRGKLLGFAIKQEIE 58

Query: 83  TDRVEEFFETKQFPDYYVKSASRVYDTEANLSVDNDLSIFPVETKENFQDGITTIAPIYG 142
            +R+++  E +QFP+ Y      +V +T ANL ++++ + FPVE KE F+ G+TTI PI G
Sbjct: 59  NERMKKMLEDRQFPEEYTTGLFKVEETSANLDINSEFTAFPVENKELFKTGLTTIVPISG 118

Query: 143 GGMRLGTFIIWRNDKEFSDDDLILVEIASTVVGIQLLNLQTENLEENIRKQTAVTMAINT 202
           GG RLGT I+ R +  F+DDDLIL E  +TVVG+++L+ +T+ +EE  R +  V MAI++
Sbjct: 119 GGQRLGTLILARLNDSFNDDDLILAEYGATVVGMEILHEKTQEIEEEARSKAVVQMAISS 178

Query: 203 LSYSEMKAVAAILGELDGLEGRLTASVIADRIGITRSVIVNALRKLESAGIIESRSLGMK 262
           LSYSE++AV  I  ELDG EG L AS IADR+GITRSVIVNALRKLESAG+IESRSLGMK
Sbjct: 179 LSYSELEAVEHIFEELDGKEGLLVASKIADRVGITRSVIVNALRKLESAGVIESRSLGMK 238

Query: 263 GTYLKVINEGIFDKLKE 279
           GTY+KV+N+    +L++
Sbjct: 239 GTYIKVLNDKFLVELEK 255
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5557> which encodes the amino acid sequence <SEQ ID 5558>. Analysis of this protein sequence reveals the following:

```
        Possible site: 38
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.64    Transmembrane  144 - 160 ( 143 - 161)

----- Final Results -----
```

```
     bacterial membrane --- Certainty=0.1256(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB13490 GB:Z99112 transcriptional regulator [Bacillus subtilis]
 Identities = 131/255 (51%), Positives = 179/255 (69%), Gaps = 3/255 (1%)

Query: 4   LLEKTRKITSILQRSVDSLETELPYNTMASRLADIIDCNACIINGGGTLLGYAMKYKTNT 63
           LL+KTR I S+LQ +       + +  MA  L D+ID N  +++  G LLGY++  +
Sbjct: 3   LLQKTRIINSMLQAAAGK---PVNFKEMAETLRDVIDSNIFVVSRRGKLLGYSINQQIEN 59

Query: 64  DRVEEFFEAKQFPDTYVKAASRVYDTEANLSVENELTIFPVESKDTYPGGLTTIAPIYGG 123
           DR+++  E +QFP+ Y K    V +T +NL + +E T FPVE++D +  GLTTI PI GG
Sbjct: 60  DRMKKMLEDRQFPEEYTKNLFNVPETSSNLDINSEYTAFPVENRDLFQAGLTTIVPIIGG 119

Query: 124 GMRLGSLIIWRNDNEFSDDDLILVEISSTVVGIQLLNLQTENLEDTIRKQTAVNMAINTL 183
           G RLG+LI+ R  ++F+DDDLIL E  +TVVG+++L  + E +E+  R +  V MAI++L
Sbjct: 120 GERLGTLILSRLQDQFNDDDLILAEYGATVVGMEILREKAEEIEEEARSKAVVQMAISSL 179

Query: 184 SYSEMKAVAAILGELDGNEGRLTASVIADRIGITRSVIVNALRKLESAGIIESRSLGMKG 243
           SYSE++A+  I  ELDGNEG L AS IADR+GITRSVIVNALRKLESAG+IESRSLGMKG
Sbjct: 180 SYSELEAIEHIFEELDGNEGLLVASKIADRVGITRSVIVNALRKLESAGVIESRSLGMKG 239

Query: 244 TYLKVINEGIFAKLK 258
           TY+KV+N     +L+
Sbjct: 240 TYIKVLNNKFLIELE 254
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 232/260 (89%), Positives = 247/260 (94%)

Query: 21  MPNLLEKTRKITSILQRSVDSLDAELPYNTMAAQLADIIDCNACIINGGGNLLGYAMKYK 80
           MPNLLEKTRKITSILQRSVDSL+ ELPYNTMA++LADIIDCNACIINGGG LLGYAMKYK
Sbjct: 1   MPNLLEKTRKITSILQRSVDSLETELPYNTMASRLADIIDCNACIINGGGTLLGYAMKYK 60

Query: 81  TNTDRVEEFFETKQFPDYYVKSASRVYDTEANLSVDNDLSIFPVETKENFQDGITTIAPI 140
           TNTDRVEEFFE KQFPD YVK+ASRVYDTEANLSV+N+L+IFPVE+K+ +  G+TTIAPI
Sbjct: 61  TNTDRVEEFFEAKQFPDTYVKAASRVYDTEANLSVENELTIFPVESKDTYPGGLTTIAPI 120

Query: 141 YGGGMRLGTFIIWRNDKEFSDDDLILVEIASTVVGIQLLNLQTENLEENIRKQTAVTMAI 200
           YGGGMRLG+ IIWRND EFSDDDLILVEI+STVVGIQLLNLQTENLE+ IRKQTAV MAI
Sbjct: 121 YGGGMRLGSLIIWRNDNEFSDDDLILVEISSTVVGIQLLNLQTENLEDTIRKQTAVNMAI 180

Query: 201 NTLSYSEMKAVAAILGELDGLEGRLTASVIADRIGITRSVIVNALRKLESAGIIESRSLG 260
           NTLSYSEMKAVAAILGELDG EGRLTASVIADRIGITRSVIVNALRKLESAGIIESRSLG
Sbjct: 181 NTLSYSEMKAVAAILGELDGNEGRLTASVIADRIGITRSVIVNALRKLESAGIIESRSLG 240

Query: 261 MKGTYLKVINEGIFDKLKEY 280
           MKGTYLKVINEGIF KLKE+
Sbjct: 241 MKGTYLKVINEGIFAKLKEF 260
```

A related GBS gene <SEQ ID 8901> and protein <SEQ ID 8902> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 9
McG: Discrim Score:     -6.84
GvH: Signal Score (-7.5): -5.37
     Possible site: 13
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value:  -2.02 threshold:  0.0
   INTEGRAL    Likelihood = -2.02   Transmembrane  114 - 130 ( 113 - 131)
   PERIPHERAL  Likelihood =  3.61     179
 modified ALOM score:   0.90

*** Reasoning Step: 3
```

```
----- Final Results -----
               bacterial membrane --- Certainty=0.1808(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF02556(223 - 987 of 1293)
EGAD|13275|BS1617(4 - 255 of 259) cody protein {Bacillus subtilis} OMNI|NT01BS1895 cody
protein (vegetative protein 286b) (veg286b) GP|535351|gb|AAB03372.1||U13634 CodY {Bacillus
subtilis} GP|2633989|emb|CAB13490.1||Z99112 transcriptional regulator {Bacillus subtilis}
PIR|S61496|S61496 transcription pleiotropic repressor codY - Bacillus subtilis
%Match = 29.1
%Identity = 50.6  %Similarity = 71.5
Matches = 128  Mismatches = 71  Conservative Sub.s = 53

177       207       237       267       297       327       357       387
DCKS*NALI*L*RKTYKG*RKCRIYLEKTRKITSILQRSVDSLDAELPYNTMAAQLADIIDCNACIINGGGNLLGYAMKY
                         |:||| | |:|| :       : :  ||  | |:|| |  :::  | ||||::
                      MALLQKTRIINSMLQAAAGK---PVNFKEMAETLRDVIDSNIFVVSRRGKLLGYSINQ
                              10        20           30        40        50

417       447       477       507       537       567       597       627
KTNTDRVEEFFETKQFPDYYVKSASRVYDTEANLSVDNDLSIFPVETKENFQDGITTIAPIYGGGMRLGTFIIWRNDKEF
:   ||::: :| :|||: | |:   | :| :|| :::: : |||| :: || |:||| || ||| ||||:|: |   :|
QIENDRMKKMLEDRQFPEEYTKNLFNVPETSSNLDINSEYTAFPVENRDLFQAGLTTIVPIIGGGERLGTLILSRLQDQF
             70        80        90       100       110       120       130

657       687       717       747       777       807       837       867
SDDDLILVEIASTVVGIQLLNLQTENLEENIRKQTAVTMAINTLSYSEMKAVAAILGELDGLEGRLTASVIADRIGITRS
:|||||| |  :||||:::|  : | :||  | :  | |||::|||||::|:  |: |||| || | || ||||:|||||
NDDDLILAEYGATVVGMEILREKAEEIEEEARSKAVVQMAISSLSYSELEAIEHIFEELDGNEGLLVASKIADRVGITRS
            150       160       170       180       190       200       210

897       927       957       987      1017      1047      1077      1107
VIVNALRKLESAGIIESRSLGMKGTYLKVINEGIFDKLKEYN*S*HGTGSSFQFLFWNQEEIRRKMTXXN*LXXLFS*RL
|||||||||||||:||||||||||||:||:|   : :|:
VIVNALRKLESAGVIESRSLGMKGTYIKVLNNKFLIELENLKSH
            230       240       250
```

SEQ ID 8902 (GBS431) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 173 (lane 7; MW 54kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 6; MW 29kDa).

GBS431-GST was purified as shown in Figure 223, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1787

A DNA sequence (GBSx1894) was identified in *S.agalactiae* <SEQ ID 5559> which encodes the amino acid sequence <SEQ ID 5560>. This protein is predicted to be isochorismatase. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.81   Transmembrane  126 - 142 ( 125 - 142)

----- Final Results -----
               bacterial membrane --- Certainty=0.2126(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15164 GB:Z99120 similar to pyrazinamidase/nicotinamidase
          [Bacillus subtilis]
 Identities = 99/181 (54%), Positives = 132/181 (72%)

Query: 1   MTKALISIDYTYDFVADDGKLTAGKPAQSIASAIADVTEKAYRSGDYIFFAIDNHDIGDV 60
           M KALI IDYT DFVA DGKLT G+P + I  AI ++T++   +GDY+  A+D+HD GD
Sbjct: 1   MKKALICIDYTNDFVASDGKLTCGEPGRMIEEAIVNLTKEFITNGDYVVLAVDSHDEGDQ 60

Query: 61  FHPESNLFPEHNIKGTSGRNLYGPLGTLYETIKEDSRVFWIDKRHYSAFSGTDLDIRLRE 120
           +HPE+ LFP HNIKGT G++LYG L  LY+  + +  V++++K  YSAF+GTDL+++LRE
Sbjct: 61  YHPETRLFPPHNIKGTEGKDLYGKLLPLYQKHEHEPNVYYMEKTRYSAFAGTDLELKLRE 120

Query: 121 RRVDTLILTGVLTDICVLHTAIDAYNLGYKIEVPAAAVASLNDSNHQWALNHFKTVLGATI 181
           R++  L L GV TDICVLHTA+DAYN G++I V   AVAS N   H WAL+HF   +GA +
Sbjct: 121 RQIGELHLAGVCTDICVLHTAVDAYNKGFRIVVHKQAVASFNQEGHAWALSHFANSIGAQV 181
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5561> which encodes the amino acid sequence <SEQ ID 5562>. Analysis of this protein sequence reveals the following:

```
      Possible site: 31
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.60   Transmembrane  126 - 142 ( 126 - 142)

----- Final Results -----
              bacterial membrane --- Certainty=0.2041(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB15164 GB:Z99120 similar to pyrazinamidase/nicotinamidase
          [Bacillus subtilis]
 Identities = 90/179 (50%), Positives = 127/179 (70%)

Query: 3   RALISIDYTNDFVADDGKLSAGKSAQAIATKIAEVTKTAFDQGDYIFFAIDCHDQNDSWH 62
           +ALI IDYTNDFVA DGKL+ G+  + I   I  +TK     GDY+  A+D HD+ D +H
Sbjct: 3   KALICIDYTNDFVASDGKLTCGEPGRMIEEAIVNLTKEFITNGDYVVLAVDSHDEGDQYH 62

Query: 63  PESKLFAAHNIKGTTGRHLYGPLAEVYSYMKQHPRVFWIDKRYYSAFSGTDLDIRLRERG 122
           PE++LF  HNIKGT G+ LYG L  +Y   +  P V++++K  YSAF+GTDL+++LRER
Sbjct: 63  PETRLFPPHNIKGTEGKDLYGKLLPLYQKHEHEPNVYYMEKTRYSAFAGTDLELKLRERQ 122

Query: 123 ITQLVLTGVLSDICVLHTAIDAYHLGYQLEIVKSAVASLTKESYEWSLAHFEQVLGAKL 181
           I +L L GV +DICVLHTA+DAY+ G+++ + K AVAS  +E + W+L+HF   +GA++
Sbjct: 123 IGELHLAGVCTDICVLHTAVDAYNKGFRIVVHKQAVASFNQEGHAWALSHFANSIGAQV 181
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 121/180 (67%), Positives = 150/180 (83%)

Query: 3   KALISIDYTYDFVADDGKLTAGKPAQSIASAIADVTEKAYRSGDYIFFAIDNHDIGDVFH 62
           +ALISIDYT DFVADDGKL+AGK AQ+IA+ IA+VT+ A+  GDYIFFAID HD  D +H
Sbjct: 3   RALISIDYTNDFVADDGKLSAGKSAQAIATKIAEVTKTAFDQGDYIFFAIDCHDQNDSWH 62

Query: 63  PESNLFPEHNIKGTSGRNLYGPLGTLYETIKEDSRVFWIDKRHYSAFSGTDLDIRLRERR 122
           PES LF  HNIKGT+GR+LYGPL  +Y  +K+  RVFWIDKR+YSAFSGTDLDIRLRER
Sbjct: 63  PESKLFAAHNIKGTTGRHLYGPLAEVYSYMKQHPRVFWIDKRYYSAFSGTDLDIRLRERG 122

Query: 123 VDTLILTGVLTDICVLHTAIDAYNLGYKIEVPAAAVASLNDSNHQWALNHFKTVLGATIL 182
           +  L+LTGVL+DICVLHTAIDAY+LGY++E+  +AVASL   +++W+L HF+ VLGA ++
Sbjct: 123 ITQLVLTGVLSDICVLHTAIDAYHLGYQLEIVKSAVASLTKESYEWSLAHFEQVLGAKLI 182
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1788

A DNA sequence (GBSx1895) was identified in *S.agalactiae* <SEQ ID 5563> which encodes the amino acid sequence <SEQ ID 5564>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.1539(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1789

A DNA sequence (GBSx1896) was identified in *S.agalactiae* <SEQ ID 5565> which encodes the amino acid sequence <SEQ ID 5566>. This protein is predicted to be 3-hydroxyacyl-CoA dehydrogenase (hbd-10). Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -0.27   Transmembrane    3 -  19 (   1 -  19)
   INTEGRAL    Likelihood = -0.11   Transmembrane  277 - 293 ( 277 - 294)

----- Final Results -----
                bacterial membrane --- Certainty=0.1107(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF12219 GB:AE001862 3-hydroxyacyl-CoA dehydrogenase, putative
          [Deinococcus radiodurans]
 Identities = 151/321 (47%), Positives = 196/321 (61%), Gaps = 36/321 (11%)

Query: 56  NMTIKNLTVAGSGVLGSQIAFQAAYKGMSVTIYDINDEALNKGKERIKKLAKVYQSEIET 115
            +M+IK +TV GSGVLGSQIAFQ A+ G  V +YDIND A+ K +E + KL   YQ +++
Sbjct: 51  SMSIKTVTVCGSGVLGSQIAFQTAFHGFDVHLYDINDAAIAKARETLGKLQARYQQDLKV 110

Query: 116 AKEAYSDKAKSIKYNKNLLPSLDHIFLSKVADSLDLIADLPNQITFSKNLDQAVSDADLV 175
              +   D                                +I+F  ++ +AV   DLV
Sbjct: 111 DAQQTGDAFA--------------------------------RISFFTDIAEAVKGVDLV 138

Query: 176 IEAVPETVSIKEDFYKQLAKVAPSKTIFATNSSTLVPSQFADITGRPDKFLAMHFANNIW 235
            IEA+PE + IK  FY QL +VA   TIFATNSSTL+PSQF + TGRP+KFLA+HFAN IW
Sbjct: 139 IEAIPENMDIKRKFYNQLGEVADPNTIFATNSSTLLPSQFMEETGRPEKFLALHFANEIW 198

Query: 236 QNNIVEIMGHKGTDDEVIKEALAFSKDIGMVPLHIHKEQPGYILNSILVPFLESALALYY 295
            + N  EIM    TDD V    + F+KDIGMV L ++KEQ GYILN++LVP L +AL L
Sbjct: 199 KFNTAEIMRTPRTDDAVFDTVVQFAKDIGMVALPMYKEQAGYILNTLLVPLLGAALELVV 258

Query: 296 DKVSDSETIDKTWKLGTGAPMGPLEILDIIGIDTAYNIMKNYSDTNSDPNSLHAHLAKML 355
              ++D +T+DKTW + TGAP GP   LD+IG+ T YNI  N +   ++P S  A  AK +
Sbjct: 259 KGIADPQTVDKTWMIATGAPRGPFAFLDVIGLTTPYNI--NMASAETNPGS--AAAAKYI 314

Query: 356 KEEFIDKGRTGKAAGHGFYDY 376
            KE +IDKG+ G A G GFY Y
```

```
Sbjct: 315 KENYIDKGKLGTATGEGFYKY 335
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8903> and protein <SEQ ID 8904> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 3
SRCFLG: 0
McG: Length  of  UR:   20
     Peak Value of UR:   1.55
     Net Charge of CR: 1
McG: Discrim Score:     -0.60
GvH: Signal Score (-7.5): -3.93
     Possible site: 21
>>> Seems to have no N-terminal signal sequence
Amino Acid Composition: calculated from 1
ALOM program   count: 1 value:  -0.11 threshold:  0.0
   INTEGRAL    Likelihood = -0.11   Transmembrane  221 - 237 ( 221 - 238)
   PERIPHERAL  Likelihood =  4.61       6
 modified ALOM score:   0.52
icml HYPID: 7  CFP: 0.104

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.1044(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
37.5/60.5% over 278aa
                                                                Archaeoglobus
fulgidus
  EGAD|103851|  3-hydroxyacyl-CoA  dehydrogenase  Insert  characterized  OMNI|AF2273  3-
hydroxyacyl-CoA dehydrogenase (hbd-10) Insert
 characterized
  GP|2648250|gb|AAB88983.1||AE000948   3-hydroxyacyl-CoA   dehydrogenase   (hbd-10)   Insert
characterized
  PIR|A69534|A69534 3-hydroxyacyl-CoA dehydrogenase (hbd-10) homolog - Insert characterized

ORF01176(475 - 1431 of 1731)
EGAD|103851|AF2273(17  -  295  of  668)  3-hydroxyacyl-CoA  dehydrogenase  {Archaeoglobus
fulgidus}OMNI|AF2273          3-hydroxyacyl-CoA            dehydrogenase             (hbd-
10)GP|2648250|gb|AAB88983.1||AE000948     3-hydroxyacyl-CoA     dehydrogenase     (hbd-10)
{Archaeoglobus fulgidus}PIR|A69534|A69534 3-hydroxyacyl-CoA dehydrogenase (hbd-10) homolog
- Archaeoglobus fulgidus
%Match = 14.8
%Identity = 37.5  %Similarity = 60.4
Matches = 106  Mismatches = 106  Conservative Sub.s = 65

387       417       447       477       507       537       567       597
KKRYYFKNNHTIYLLLDISFVKLSSKTFSNISIGGCNMTIKNLTVAGSGVLGSQIAFQAAYKGMSVTIYDINDEALNKGK
                         :    :        || : |:|::|  ||   |  | :||: ||  | :::|
                       MPRRVKQVINMDVRERIKTVAVLGAGLMGHGIAEVCAMAGYNVTMRDIKQEFVDRGM
                               10        20        30        40        50

624       651       681       711       741       771       801       831
ERIKK-LAKVYQS-EIETAKEAYSDKAKSIKYNKNLLPSLDHIFLSKVADSLDLIADLPNQITFSKNLDQAVSDADLVIE
  ||: |||: |  :|::|:|  |                                    :|  : :|::|| |||||||
NMIKESLAKLEQKGKIKSAEEVLS------------------------------------RIKPTVDLEEAVKDADLVIE
           70        80                                             90       100
```

```
861       891       921       951       981      1011      1041      1071
AVPETVSIKEDFYKQLAKVAPSKTIFATNSSTLVPSQFADITGRPDKFLAMHFANNIWQNNIVEIMGHKGTDDEVIKEAL
|||| | ||:  :::: |:|    || :|:||:  : :|| | ||:||  :|| |      :||::  : | |||:   :
AVPEVVEIKKQVWEEVDKLAKPDCIFTSNTSTMRITMLADFTSRPEKFAGLHFFNPPVLMRLVEVIRGEKTSDEVMDLLV
                120       130       140       150       160       170       180

1101      1131      1161      1191      1221      1251      1281      1311
AFSKDIGMVPLHIHKEQPGYILNSILVPFLESALALYYDKVSDSETIDKTWKLGTGAPMGPLEILDIIGIDTAYNIMKNY
 | | ||  |: : |: ||:|:| :  |     :|:    ::  | :| | :   | ||||:|::|  |:|  || :| |
EFVKSIGKTPVRVEKDVPGFIVNRVQAPASVLLMAILEKGIATPEEVDATVR-RLGLPMGPFELVDYTGVDILYNALKYY
                200       210       220       230        240       250        260

1341      1371      1401      1431      1461      1491      1521      1551
SDTNSDPNSLHAHLAKMLKEEFIDKGRTGKAAGHGFYDYD*TIKEVR*KSNLFYNSTKE*LHQEQF*NDLKPIDDYYHLS
: | | |:    :    : || :   : |:  |:||||:     ::         :  :    |
AQTIS-PD----YEPPKFLEEMVKANKLGRKTGQGFYDWSKGRPQIDSSKATDKINPMDFTFVEINEAVKLVEMGVATPQ
            270       280       290       300       310       320       330
```

SEQ ID 8904 (GBS112) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 28 (lane 5; MW 39kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 34 (lane 3; MW 64kDa).

GBS112-GST was purified as shown in Figure 198, lane 10.

Example 1790

A DNA sequence (GBSx1897) was identified in *S.agalactiae* <SEQ ID 5567> which encodes the amino acid sequence <SEQ ID 5568>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3332(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10097> which encodes amino acid sequence <SEQ ID 10098> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14467 GB:Z99117 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 62/169 (36%), Positives = 109/169 (63%), Gaps = 3/169 (1%)

Query: 1    MAVLSMLGIIDAKPKVGYFYLGQYHASIGTSHFEKMTVSEIMGIPLTVHQKDSVYDVIVH 60
            +A+L+M G ++A+P+VGYFY G+    +     +K+ V +   IP+ +H+  SVYD I
Sbjct: 43   LAILTMSGFLEARPRVGYFYTGKTGTQLLADKLKKLQVKDFQSIPVVIHENVSVYDAICT 102

Query: 61   IFMEDAGCAFILDDDDFLCGVVSRKDLLKISIGGGDLSKMPIGMVMTRMPHVTTVLENES 120
            +F+ED G  F++D D  L GV+SRKDLL+ SIG  +L+ +P+ ++MTRMP++T     +
Sbjct: 103  MFLEDVGTLFVVDRDAVLVGVLSRKDLLRASIGQQELTSVPVHIIMTRMPNITVCRREDY 162

Query: 121  LFAAADKLVSRKVDSLPVVRHDKQYPEKFKVIGKLSKTILASLFLEIRD 169
            +   A  L+ +++D+LPV+   K   + F+VIG+++KT +  + + + +
Sbjct: 163  VMDIAKHLIEKQIDALPVI---KDTDKGFEVIGRVTKTNMTKILVSLSE 208
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1791

A DNA sequence (GBSx1898) was identified in *S.agalactiae* <SEQ ID 5569> which encodes the amino acid sequence <SEQ ID 5570>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.53   Transmembrane   60 -  76 (  60 -  76)

----- Final Results -----
               bacterial membrane --- Certainty=0.1213(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB05092 GB:AP001511 unknown conserved protein [Bacillus halodurans]
 Identities = 126/256 (49%), Positives = 183/256 (71%), Gaps = 1/256 (0%)

Query: 7   IFIISDSLGETAKAIAKACLSQFPGHDDWHFQRFSYINSQERLEQVFEEASQKTVFMMFS 66
            ++++SDS+GETA+ + KA  SQF G      +R  Y+  +E +++V + A Q    + F+
Sbjct: 10  VYVVSDSVGETAELVVKAAASQFSGAGI-EVRRIPYVEDKETVDEVIQLAKQADAIIAFT 68

Query: 67  LVDVALASYAQKRCESEHYAYVDLLTNVIQGISRISGIDPLGEPGILRRLDNDYFKRVES 126
            LV   + +Y  ++        VD++  +++ IS ++  +P  EPGI+ RLD DYF++VE+
Sbjct: 69  LVVPGIRTYLLEKATEAKVETVDIIGPMLEKISSLTKEEPRYEPGIVYRLDEDYFRKVEA 128

Query: 127 IEFAVKYDDGRDPRGILQADLVIIGISRTSKTPLSMFLADKNIKVINIPLVPEVPVPKEL 186
            IEFAVKYDDGRDPRGI++ADLV+IG+SRTSKTPLS +LA K +KV N+PLVPEV  P+EL
Sbjct: 129 IEFAVKYDDGRDPRGIVRADLVLIGVSRTSKTPLSQYLAHKRLKVANVPLVPEVEPPEEL 188

Query: 187 RMIDSRRIIGLTNSVDHLNQVRKVRLKSLGLSSTANYASLERILEETRYAEEVMKNLGCP 246
              +  +++IGL  S + LN +R  RLK+LGL S ANYA+++RI EE  YAE +MK +GCP
Sbjct: 189 FKLSPKKVIGLKISPEQLNGIRAERLKTLGLKSQANYANIDRIKEELAYAEGIMKRIGCP 248

Query: 247 IINVSDKAIEETATII 262
            +I+VS+KA+EETA +I
Sbjct: 249 VIDVSNKAVEETANLI 264
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 5570 (GBS378) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 68 (lane 4; MW 34kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 2; MW 59kDa).

GBS378-GST was purified as shown in Figure 212, lane 6.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1792

A DNA sequence (GBSx1899) was identified in *S.agalactiae* <SEQ ID 5571> which encodes the amino acid sequence <SEQ ID 5572>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3703(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD35361 GB:AE001709 pyruvate,orthophosphate dikinase
            [Thermotoga maritima]
 Identities = 494/882 (56%), Positives = 639/882 (72%), Gaps = 9/882 (1%)

Query: 1    METKFVYHFD----EGCKEMKKELLGGKGANLAEMTSIGLPVPQGFTITTQACNDYYDNAC 56
            M  K+VY F     EG  +MK++LGGKGANLAEMT++G+PVP GFTI+ + C  YYD+
Sbjct: 1    MAKKYVYFFANGKAEGRADMKDILGGKGANLAEMTNLGIPVPPGFTISAEVCKYYYDHGR 60

Query: 57   HIRESILSQIDQAMAQLEVEQNKQLGSVDDPLLVSVRSGSVFSMPGMMDTVLNLGLNDRS 116
               E +  Q+++AM +LE    K+ G  ++PLLVSVRSG+  SMPGMMDTVLNLGLND +
Sbjct: 61   TYPEELKEQVEEAMRRLEEVTGKKFGDPNNPLLVSVRSGAAISMPGMMDTVLNLGLNDET 120

Query: 117  VQGLVKKTEDERFAYDSYRRFIQMFADVVTGIPKYKFDTILDRLKTDKCYQDDTELTGSD 176
            V+GL K T +ERFAYD+YRRF+QMF DVV  IP  KF+  L+ LK +K  + DTEL   D
Sbjct: 121  VKGLAKLTNNERFAYDAYRRFLQMFGDVVLKIPHEKFEKALEELKKEKGVKLDTELDAED 180

Query: 177  LKRLVEFYKELYQKEAGEKFPQDPKRQLLLAIEAVFKSWNNPRAKIYRKLNDIPE--TLG 234
            LK+LVE YK++Y KE G++FPQDP +QL LAI+AVF SW N RA  YR+++ I E   LG
Sbjct: 181  LKKLVERYKQIY-KEEGKEFPQDPWKQLWLAIDAVFGSWMNERAIKYRQIHGIKEGDLLG 239

Query: 235  TAVNIQAMVFGNMGNNSGTGVAFTRNPSTGAANLFGEYLINAQGEDVVAGIRTPQSISKL 294
            TAVNI AMVFGNMG +SGTGVAFTR+P+TG    +GE+L NAQGEDVVAGIRTP  + +L
Sbjct: 240  TAVNIVAMVFGNMGEDSGTGVAFTRDPNTGEKKPYGEFLPNAQGEDVVAGIRTPLKLEEL 299

Query: 295  AEQMPIIYQEFVSVTQKLEAHYRDMQDMEFTIENGNLYMLQTRSGKRTAKAAIKIAVDQV 354
              +MP +Y + + +  KLE HYRDMQD+EFT+E G LY+LQTR+GKRT++AAI+IAVD V
Sbjct: 300  KNRMPEVYNQLLEIMDKLEKHYRDMQDIEFTVERGKLYILQTRNGKRTSQAAIRIAVDMV 359

Query: 355  NEGLISKEEAILRIEPKQLDQLLHPSFDLKSLKKAIILTTGLPASPGAAYGKVYFHAEDV 414
            +EGLI+KEEAILR+ P+ ++Q+LHP FD K   +A ++  GLPASPGAA GKV F+A+
Sbjct: 360  HEGLITKEEAILRVRPEDVEQVLHPVFDPKEKAQAKVIAKGLPASPGAATGKVVFNAKKA 419

Query: 415  VKEMKKGNPVLLVRQETSPEDIEGMVSANGIITARGGMTSHAAVVARGMGKPCVAGCSQL 474
             +  K G  V+LVR ETSPED+ GM +A GI+T+RGGMTSHAAVVARGMGKP V G   +
Sbjct: 420  EELGKAGEQVILVRPETSPEDVGGMAAAQGILTSRGGMTSHAAVVARGMGKPAVVGAESI 479

Query: 475  LVDEVRREISIGHQTIKEGEMLSIDGATGNVYIGQV-PMAETSVDRDFEIFMKWVDENRD 533
             V        +G   +KEGE +SIDG TG V +G+V  +    ++      ++W DE R
Sbjct: 480  EVHPEEGYFKVGDVVVKEGEWISIDGTTGEVLLGKVTTIKPQGLEGPVAELLQWADEIRR 539

Query: 534  MMVCSNADNPRDAQKALDFGAEGIGLCRTEHMFFDDERIPVVREMILADEILSRRKALER 593
            + V +NAD PRDA+ A  FGAEGIGLCRTEHMFF+ +RIP VR MILA     R KAL+
Sbjct: 540  LGVRTNADIPRDAEVARKFGAEGIGLCRTEHMFFEKDRIPKVRRMILAKTKEEREKALDE 599

Query: 594  LLSFQRDDFYQIFKVLKGKACTIRLLDPPLHEFLPHDKESIESMARQMGISTLAIEKRIQ 653
            LL  Q++DF  +F+V+KG   TIRL+DPPLHEFLP + E I+ +A QMG+S   ++  ++
Sbjct: 600  LLPLQKEDFKGLFRVMKGLPVTIRLIDPPLHEFLPQEDEQIKEVAEQMGVSFEELKNVVE 659

Query: 654  TLEEFNPMLGHRGCRLAITYPEIYQMQVRALVQGAI-LAMKEGYEAKPEIMIPLVTAHEE 712
             L+E NPMLGHRGCRL ITYPEI  MQ +A++  AI L  +EG +  PEIMIPLV    E
Sbjct: 660  NLKELNPMLGHRGCRLTITYPEIAVMQTKAIIGAAIELKKEEGIDVIPEIMIPLVGHVNE 719

Query: 713  ISIIRDLIEETIVEESKSKKINLSFPIGTMIETPRACMIADDIAKFADFFSFGTNDLTQM 772
            +  ++ +I+ET     K   + L++ IGTMIE PRA + A  IA+ A+FFSFGTNDLTQM
Sbjct: 720  LRYLKKIIKETADALIKEAGVELTYKIGTMIEVPRAAVTAHQIAEEAEFFSFGTNDLTQM 779

Query: 773  SFGFSRDDAGKFLGEYVDKGLLKKDPFQVLDQKGIGRFIGQAVRLGKEVKPNLKIGICGE 832
            +FGFSRDD GKFL EY++KG+L+ DPF+ LD  G+G  +      G+  +P+LK+G+CGE
Sbjct: 780  TFGFSRDDVGKFLPEYLEKGILEHDPFKTLDYDGVGELVRMGKEKGRSTRPDLKVGVCGE 839

Query: 833  HGGEPSSIEFCYQLGLHYVSCSPFRIPIAKLAAAQAKIKQSR 874
            HGG+P SI F  ++GL YVSCSP+R+P+A+LAAAQA +K  +
Sbjct: 840  HGGDPRSILFFDKIGLDYVSCSPYRVPVARLAAAQAALKNKK 881
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1793

A DNA sequence (GBSx1900) was identified in *S.agalactiae* <SEQ ID 5573> which encodes the amino acid sequence <SEQ ID 5574>. This protein is predicted to be glutamyl-tRNA (Gln) amidotransferase subunit C (gatC). Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3229(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04384 GB:AP001509 glutamyl-tRNA (Gln) amidotransferase
         subunit C [Bacillus halodurans]
 Identities = 42/94 (44%), Positives = 63/94 (66%)

Query: 2  KISEEEVRHVANLSKLRFSDQETKEFASSLSKIVDMIELLNEVDTEGVPVTTTMADRKTV 61
          +IS E+V+HVA+L++L  +++E K F   L  I+   E LNE+DTEGV  T+ + D K V
Sbjct: 3  RISMEQVKHVAHLARLAITEEEAKLFTEQLGDIIQFAEQLNELDTEGVEPTSHVLDMKNV 62

Query: 62 MREDIAQPGHNRDDLFKNVPQHQDYYIKVPAILE 95
          +RED  + G   +D+ KN P H+D  I+VP++LE
Sbjct: 63 LREDKPEKGLPVEDVLKNAPDHEDGQIRVPSVLE 96
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5575> which encodes the amino acid sequence <SEQ ID 5576>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3247(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 72/100 (72%), Positives = 88/100 (88%)

Query: 1    MKISEEEVRHVANLSKLRFSDQETKEFASSLSKIVDMIELLNEVDTEGVPVTTTMADRKT 60
            MKISEEEVRHVA LSKL FS+ ET  FA++LSKIVDM+ELLNEVDTEGV +TTTMAD+K
Sbjct: 5    MKISEEEVRHVAKLSKLSFSESETTTFATTLSKIVDMVELLNEVDTEGVAITTTMADKKN 64

Query: 61   VMREDIAQPGHNRDDLFKNVPQHQDYYIKVPAILEDGGDA 100
            VMR+D+A+ G +R  LFKNVP+ ++++IKVPAIL+DGGDA
Sbjct: 65   VMRQDVAEEGTDRALLFKNVPEKENHFIKVPAILDDGGDA 104
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1794

A DNA sequence (GBSx1901) was identified in *S.agalactiae* <SEQ ID 5577> which encodes the amino acid sequence <SEQ ID 5578>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.64    Transmembrane    7 -  23 (   6 -  24)

----- Final Results -----
               bacterial membrane --- Certainty=0.4057(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1795

A DNA sequence (GBSx1902) was identified in *S.agalactiae* <SEQ ID 5579> which encodes the amino acid sequence <SEQ ID 5580>. This protein is predicted to be glutamyl-tRNA amidotransferase, subunit A (gatA). Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2855(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04385 GB:AP001509 glutamyl-tRNA (Gln) amidotransferase
           subunit A [Bacillus halodurans]
 Identities = 285/486 (58%), Positives = 367/486 (74%), Gaps = 4/486 (0%)

Query: 1   MSFNNQSIDQLHDFLVKKEISATELTKATLEDIHAREQAVGSFITISDEMAIAQAKEID- 59
           MS  +  +  +H  L +KEIS ++L     + I   +  V +F+ +++E A A AKE+D
Sbjct: 1   MSLFDLKLKDVHTKLHEKEISVSDLVDEAYKRIEQVDGQVEAFLALNEEKARAYAKELDA 60

Query: 60  --DKGIDADNVMSGIPLAVKDNISTKGILTTAASKMLYNYEPIFDATAVEKLYAKDMIVI 117
             D+  +A  ++ GIP+ VKDNI TK + TT +S++L N++PI+DAT V KL     + I
Sbjct: 61  ALDRS-EARGLLFGIPIGVKDNIVTKNLRTTCSSRILGNFDPIYDATVVHKLREAQAVTI 119

Query: 118 GKANMDEFAMGGSTETSYFKKTNNAWDHSKVPGGSSGGSAAAVASGQVRLSLGSDTGGSI 177
           GK NMDEFAMG STE S F+KT N W+   VPGGSSGGSAAAVA+G+V  +LGSDTGGSI
Sbjct: 120 GKLNMDEFAMGSSTENSAFQKTKNPWNLEYVPGGSSGGSAAAVAAGEVPFTLGSDTGGSI 179

Query: 178 RQPASFNGIVGMKPTYGRVSRFGLFAFGSSLDQIGPMSQTVKENAQLLTVISGHDVRDST 237
           RQPA++ G+VG+KPTYGRVSR+GL AF SSLDQIGP+++ V++NA LL  ISGHD  DST
Sbjct: 180 RQPAAYCGVVGLKPTYGRVSRYGLVAFASSLDQIGPITRNVEDNAYLLQAISGHDPMDST 239

Query: 238 SSERTVGDFTAKIGQDIQGMKIALPKEYLGEGIAQGVKETIIKAAKHLEKLGAVIEEVSL 297
           S+   V D+ + +  DI+G+KIA+PKEYLGEG+ + VK++++ A K LE LGA  EEVSL
Sbjct: 240 SANLDVPDYLSALTGDIKGLKIAVPKEYLGEGVKEEVKQSVLDALKVLEGLGATWEEVSL 299

Query: 298 PHSKYGVAVYYIVASSEASSNLQRFDGIRYGYRTENYKNLDDIYVNTRSEGFGDEVKRRI 357
           PHSKY +A YY++ASSEAS+NL RFDG+RYG+R++N  NL D+Y  TR+EGFGDEVKRRI
Sbjct: 300 PHSKYALATYYLLASSEASANLARFDGVRYGFRSDNADNLLDMYKQTRAEGFGDEVKRRI 359

Query: 358 MLGTFSLSSGYYDAYYKKAGQVRSLIIQDFEKVFADYDLILGPTAPTTAFDLDSLNHDPV 417
           MLGTF+LSSGYYDAYYKKA QVR+LI QDFEKVF  YD+I+GPT PT AF +     DP+
Sbjct: 360 MLGTFALSSGYYDAYYKKAQQVRTLIKQDFEKVFEQYDVIIGPTTPTPAFKIGEKTDDPL 419

Query: 418 AMYLADILTIPVNLAGLPGISIPAGFDQGLPVGMQLIGPKFSEETIYQVAAAFEATTDYH 477
```

```
                MY  DILTIPVNLAG+P IS+P GFD GLP+G+Q+IG  F E ++Y+VA AFE  TDYH
Sbjct: 420 TMYANDILTIPVNLAGVPAISVPCGFDNGLPLGLQIIGKHFDEGSVYRVAHAFEQATDYH 479

Query: 478 KQQPKI 483
            ++P +
Sbjct: 480 TKRPTL 485
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5581> which encodes the amino acid sequence <SEQ ID 5582>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2364(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 392/487 (80%), Positives = 442/487 (90%)

Query: 1   MSFNNQSIDQLHDFLVKKEISATELTKATLEDIHAREQAVGSFITISDEMAIAQAKEIDD 60
           MSFN+++I++LHD LV KEISATELT+ATLEDI +RE+AVGSFIT+S+E+A+ QA  ID
Sbjct: 1   MSFNHKTIEELHDLLVAKEISATELTQATLEDIKSREEAVGSFITVSEEVALKQAAAIDA 60

Query: 61  KGIDADNVMSGIPLAVKDNISTKGILTTAASKMLYNYEPIFDATAVEKLYAKDMIVIGKA 120
           KGIDADN+MSGIPLAVKDNISTK ILTTAASKMLYNYEPIF+AT+V   YAKDMIVIGK
Sbjct: 61  KGIDADNLMSGIPLAVKDNISTKEILTTAASKMLYNYEPIFNATSVANAYAKDMIVIGKT 120

Query: 121 NMDEFAMGGSTETSYFKKTNNAWDHSKVPGGSSGGSAAAVASGQVRLSLGSDTGGSIRQP 180
           NMDEFAMGGSTETSYFKKT NAWDH+KVPGGSSGGSA AVASGQVRLSLGSDTGGSIRQP
Sbjct: 121 NMDEFAMGGSTETSYFKKTKNAWDHTKVPGGSSGGSATAVASGQVRLSLGSDTGGSIRQP 180

Query: 181 ASFNGIVGMKPTYGRVSRFGLFAFGSSLDQIGPMSQTVKENAQLLTVISGHDVRDSTSSE 240
           A+FN +VG+KPTYG VSR+GL AFGSSLDQIGP + TVKENAQLL VI+  DV+D+TS+
Sbjct: 181 AAFNSVVGLKPTYGTVSRYGLIAFGSSLDQIGPFAPTVKENAQLLNVIASSDVKDATSAP 240

Query: 241 RTVGDFTAKIGQDIQGMKIALPKEYLGEGIAQGVKETIIKAAKHLEKLGAVIEEVSLPHS 300
             + D+T+KIG+DI+GMKIALPKEYLGEGI   +KET++ + K  E LGA +EEVSLPHS
Sbjct: 241 VRIADYTSKIGRDIKGMKIALPKEYLGEGIDPEIKETVLASVKQFEALGATVEEVSLPHS 300

Query: 301 KYGVAVYYIVASSEASSNLQRFDGIRYGYRTENYKNLDDIYVNTRSEGFGDEVKRRIMLG 360
           KYGVAVYYI+ASSEASSNLQRFDGIRYG+R ++ KNLD+IYVNTRS+GFGDEVKRRIMLG
Sbjct: 301 KYGVAVYYIIASSEASSNLQRFDGIRYGFRADDAKNLDEIYVNTRSQGFGDEVKRRIMLG 360

Query: 361 TFSLSSGYYDAYYKKAGQVRSLIIQDFEKVFADYDLILGPTAPTTAFDLDSLNHDPVAMY 420
           TFSLSSGYYDAY+KKAGQVR+LIIQDF+KVFADYDLILGPT PT AF LD+LNHDPVAMY
Sbjct: 361 TFSLSSGYYDAYFKKAGQVRTLIIQDFDKVFADYDLILGPTTPTVAFGLDTLNHDPVAMY 420

Query: 421 LADILTIPVNLAGLPGISIPAGFDQGLPVGMQLIGPKFSEETIYQVAAAFEATTDYHKQQ 480
           LAD+LTIPVNLAGLPGISIPAGF  GLPVG+QLIGPK++EETIYQ AAAFEA TDYHKQQ
Sbjct: 421 LADLLTIPVNLAGLPGISIPAGFVDGLPVGLQLIGPKYAEETIYQAAAAFEAVTDYHKQQ 480

Query: 481 PKIFGGE 487
           P IFGG+
Sbjct: 481 PIIFGGD 487
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1796

A DNA sequence (GBSx1903) was identified in *S.agalactiae* <SEQ ID 5583> which encodes the amino acid sequence <SEQ ID 5584>. This protein is predicted to be glutamyl-tRNAGln amidotransferase subunit B (gatB). Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3935(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10095> which encodes amino acid sequence <SEQ ID 10096> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04386 GB:AP001509 glutamyl-tRNA (Gln) amidotransferase
          subunit B [Bacillus halodurans]
 Identities = 308/476 (64%), Positives = 361/476 (75%), Gaps = 1/476 (0%)

Query: 1   MNFETVIGLEVHVELNTNSKIFSPSSAHFGQEQNANTNVIDWSFPGVLPVMNKGVIDAGI 60
           MNFETVIGLEVHVEL T SKIFS S  HFG E NANT+VID  +PGVLPV+NK  ++  +
Sbjct: 1   MNFETVIGLEVHVELKTESKIFSASPNHFGAEPNANTSVIDLGYPGVLPVLNKAAVEFAM 60

Query: 61  KAALALNMDIHQNMHFDRKNYFYPDNPKAYQISQFDEPIGYNGWIEIELEDGTRKKIRIE 120
           KAA+ALN ++  +  FDRKNYFYPDNPKAYQISQFD+PIG NGWIEIE+ DGT+KKI I
Sbjct: 61  KAAMALNCEVATDTKFDRKNYFYPDNPKAYQISQFDKPIGENGWIEIEV-DGTKKKIGIT 119

Query: 121 RAHLEEDAGKNTHGTDGYSYVDLNRQGVPLIEIVSEADMRSPEEAYAYLTALKEIIQYTG 180
           R HLEEDAGK TH  +GYS VD NRQG PLIEIVSE D+R+P+EAYAYL  LK IIQYTG
Sbjct: 120 RLHLEEDAGKLTHSGNGYSLVDFNRQGTPLIEIVSEPDIRTPQEAYAYLEKLKSIIQYTG 179

Query: 181 ISDVKMEEGSMRVDANISLRPYGQEEFGTKAELKNLNSFNNVRKGLIHEEKRQAQVLRSG 240
           +SD KMEEGS+R DANISLRP GQEEFGTK ELKNLNSFN VRKGL +EEKRQAQVL SG
Sbjct: 180 VSDCKMEEGSLRCDANISLRPVGQEEFGTKTELKNLNSFNFVRKGLEYEEKRQAQVLLSG 239

Query: 241 GQIQQETRRFDETTGETILMRVKEGSSDYRYFPEPDLPLFDISDEWIDQVRLELPEFPQE 300
           G+I QETRR+DE   +T+LMRVKEGS DYRYFPEPDL    I DEW  ++R E+PE P
Sbjct: 240 GEILQETRRYDEAANKTVLMRVKEGSDDYRYFPEPDLVALHIDDEWKARIRSEIPELPDA 299

Query: 301 RRAKYVSSFGLSSYDASQLTATKATSDFFEKAVAIGGDAKQVSNWLQGEVAQFLNSESKS 360
           R+ +YV   GL +YDA  LT TK  SDFFE+ +A G D K  SNWL GEV+ +LN+E K
Sbjct: 300 RKKRYVEELGLPAYDAMVLTLTKEMSDFFEETIAKGADPKLASNWLMGEVSGYLNAEQKE 359

Query: 361 IEEIGLTPENLVEMIGLIADGTISSKIAKKVFVHLAKNGGSAEEFVKKAGLVQISDPEVL 420
           ++E+ LTP+ L +MI LI  GTISSKIAKKVF  L + GG  EE VK  GLVQISD   L
Sbjct: 360 LDEVALTPDGLAKMIQLIEKGTISSKIAKKVFKDLIEKGGDPEEIVKAKGLVQISDEGEL 419

Query: 421 IPIIHQVFADNEAAVIDFKSGKRNADKAFTGYLMKATKGQANPQVALKLLAQELAK 476
              + +V  +N+ ++ DFK+GK  A     G +MKATKG+ANP +  KLL +E+ K
Sbjct: 420 RKYVVEVLDNNQQSIDDFKNGKDRAIGFLVGQIMKATKGKANPPMVNKLLLEEINK 475
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5585> which encodes the amino acid sequence <SEQ ID 5586>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3935(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 410/479 (85%), Positives = 447/479 (92%)

Query: 1   MNFETVIGLEVHVELNTNSKIFSPSSAHFGQEQNANTNVIDWSFPGVLPVMNKGVIDAGI 60
            MNFET+IGLEVHVELNTNSKIFSPSSAHFG++ NANTNVIDWSFPGVLPVMNKGVIDAGI
Sbjct: 1   MNFETIIGLEVHVELNTNSKIFSPSSAHFGEDPNANTNVIDWSFPGVLPVMNKGVIDAGI 60

Query: 61  KAALALNMDIHQNMHFDRKNYFYPDNPKAYQISQFDEPIGYNGWIEIELEDGTRKKIRIE 120
            KAALALNMDIH+ MHFDRKNYFYPDNPKAYQISQFDEPIGYNGWI+I+LEDG+ KKIRIE
Sbjct: 61  KAALALNMDIHKEMHFDRKNYFYPDNPKAYQISQFDEPIGYNGWIDIKLEDGSTKKIRIE 120

Query: 121 RAHLEEDAGKNTHGTDGYSYVDLNRQGVPLIEIVSEADMRSPEEAYAYLTALKEIIQYTG 180
            RAHLEEDAGKNTHGTDGYSYVDLNRQGVPLIEIVSEADMRSPEEAYAYLTALKEIIQYTG
Sbjct: 121 RAHLEEDAGKNTHGTDGYSYVDLNRQGVPLIEIVSEADMRSPEEAYAYLTALKEIIQYTG 180

Query: 181 ISDVKMEEGSMRVDANISLRPYGQEEFGTKAELKNLNSFNNVRKGLIHEEKRQAQVLRSG 240
            ISDVKMEEGSMRVDANISLRPYGQE+FGTK ELKNLNSF+NVRKGL  E +RQA++LRSG
Sbjct: 181 ISDVKMEEGSMRVDANISLRPYGQEQFGTKTELKNLNSFSNVRKGLEFEVERQAKLLRSG 240

Query: 241 GQIQQETRRFDETTGETILMRVKEGSSDYRYFPEPDLPLFDISDEWIDQVRLELPEFPQE 300
            G I+QETRR+DE    TILMRVKEG++DYRYFPEPDLPL++I D WID++R +LP+FP +
Sbjct: 241 GVIRQETRRYDEANKGTILMRVKEGAADYRYFPEPDLPLYEIDDAWIDEMRAQLPQFPAQ 300

Query: 301 RRAKYVSSFGLSSYDASQLTATKATSDFFEKAVAIGGDAKQVSNWLQGEVAQFLNSESKS 360
            RRAKY    GLS+YDASQLTATK  SDFFE AV++GGDAKQVSNWLQGEVAQFLN+E K+
Sbjct: 301 RRAKYEEELGLSAYDASQLTATKVLSDFFETAVSLGGDAKQVSNWLQGEVAQFLNAEGKT 360

Query: 361 IEEIGLTPENLVEMIGLIADGTISSKIAKKVFVHLAKNGGSAEEFVKKAGLVQISDPEVL 420
            IEEI LTPENLVEMI +IADGTISSK+AKKVFVHLAKNGGSA  +V+KAGLVQISDP VL
Sbjct: 361 IEEIALTPENLVEMIAIIADGTISSKMAKKVFVHLAKNGGSARAYVEKAGLVQISDPAVL 420

Query: 421 IPIIHQVFADNEAAVIDFKSGKRNADKAFTGYLMKATKGQANPQVALKLLAQELAKLKE 479
            +PIIHQVFADNEAAV DFKSGKRNADKAFTG+LMKATKGQANPQVA +LLAQEL KL++
Sbjct: 421 VPIIHQVFADNEAAVADFKSGKRNADKAFTGFLMKATKGQANPQVAQQLLAQELQKLRD 479
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1797

A DNA sequence (GBSx1904) was identified in *S.agalactiae* <SEQ ID 5587> which encodes the amino acid sequence <SEQ ID 5588>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.27    Transmembrane  108 - 124 ( 105 - 125)
   INTEGRAL    Likelihood = -7.27    Transmembrane  278 - 294 ( 268 - 301)
   INTEGRAL    Likelihood = -6.05    Transmembrane  191 - 207 ( 188 - 208)
   INTEGRAL    Likelihood = -5.63    Transmembrane  219 - 235 ( 215 - 242)
   INTEGRAL    Likelihood = -3.93    Transmembrane   41 -  57 (  39 -  58)
   INTEGRAL    Likelihood = -3.88    Transmembrane  132 - 148 ( 131 - 150)
   INTEGRAL    Likelihood = -3.03    Transmembrane  254 - 270 ( 253 - 272)
   INTEGRAL    Likelihood = -3.03    Transmembrane   79 -  95 (  79 -  95)

----- Final Results -----
               bacterial membrane --- Certainty=0.3909(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10093> which encodes amino acid sequence <SEQ ID 10094> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA04271 GB:AJ000733 hypothetical protein [Bacillus megaterium]
 Identities = 102/292 (34%), Positives = 169/292 (56%), Gaps = 3/292 (1%)

Query: 6   TKKEKGTMMTLAAGLAWGISGISGQYLMSH-GVHVNLLTSLRLLITGIFLLSLARSKQKE 64
           +++  G ++ +     WG+SG   QYL  H   +   L  +R+L++G+ LL++A SKQ+
Sbjct: 1   SRRAWGLLLVIIGATMWGVSGTVAQYLFQHKSFNAEWLVVVRMLVSGLLLLAIA-SKQR- 58

Query: 65  HLVAAWKQPKFLKQVLLFSIFGLVLNQYAFLRAIHLTNAGTATVLQYMAPILILSIVCIL 124
           ++ A WK  +    +LLF + G++  QY + AI   NA TATVLQY +PI I+  + +
Sbjct: 59  NIFAIWKTKEERTSLLLFGVIGMLGVQYTYFAAIEAGNAATATVLQYTSPIFIIGYLAVQ 118

Query: 125 NRQRPTSFEIIAIAMAILGTYMIATHGKLGSLAITPKGLMWGLGSAITYSIYILLPVKLI 184
            R+ P   E+I++ + I GT+ +AT G    L+IT   L WG+G+A+T + Y L P +L+
Sbjct: 119 ARKWPVKVEMISVVLVIAGTFFLATSGNFNELSITGWALFWGIGAAVTSAFYTLQPKRLL 178

Query: 185 HEWGSTIVIGSGMFIGGILFSLVTKAWQYPLQINVMSILAYIGIIGIGTIFAYTFFLKGV 244
            +W S  V+G GM IGG  FS +   W    + +++S+ A + +I  GT+ A+  +L+ +
Sbjct: 179 AKWSSIEVVGWGMVIGGASFSFIHPPWHIAGEWSLLSLCAVLFVIIFGTLIAFYCYLESL 238

Query: 245 SIVGAVKGSLLASVEPVSSVFLTVLVLGEIFYPIDLLGMLFIFLAVTLISYK 296
             + A +  +LAS EP+S+  L+VL L   F   + LG + I   V L+S +
Sbjct: 239 KHISASEAIVLASREPLSAAALSVLWLHVTFGWTEWLGTILIIATVFLLSQR 290
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1798

A DNA sequence (GBSx1905) was identified in *S.agalactiae* <SEQ ID 5589> which encodes the amino acid sequence <SEQ ID 5590>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2103(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10091> which encodes amino acid sequence <SEQ ID 10092> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14510 GB:Z99117 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 52/153 (33%), Positives = 88/153 (56%)

Query: 17  YRPTFVVEAVYDLRAEDLLRHGIRAVLVDLDNTLIAWNNPDGTAEVRAWLDEMTTADISV 76
           + P   V+ ++ +  E L    ++ ++ DLDNTL+ W+ P+ T  +  W +EM    I V
Sbjct: 6   FLPDEFVKNIFHITPEKLKERNVKGIITDLDNTLVEWDRPNATPRLIEWFEEMKEHGIKV 65

Query: 77  VVVSNNNHARVERAVSRFGVDFVSRAMKPFTRGINMAIERYGFDRDEVIMVGDQLMTDIR 136
            +VSNNN  RV+      G+ F+ +A KP  +  N A+      +++ +++GDQL+TD+
Sbjct: 66  TIVSNNNERRVKLFSEPLGIPFIYKARKPMGKAFNRAVRNMELKKEDCVVIGDQLLTDVL 125

Query: 137 ASHRAGIKSVLVKPIVKSDAWNTKFNRLRERRV 169
             +R G  ++LV P+  SD + T+FNR  ERR+
Sbjct: 126 GGNRNGYHTILVVPVASSDGFITRFNRQVERRI 158
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5591> which encodes the amino acid sequence <SEQ ID 5592>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4252(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 147/175 (84%), Positives = 158/175 (90%)

Query: 12  LSIDDYRPTFVVEAVYDLRAEDLLRHGIRAVLVDLDNTLIAWNNPDGTAEVRAWLDEMTT 71
            +SIDDYRPT++VEA+YDLRA DLLRHGI AVLVDLDNTLIAWNNPDGT EVRAWLDEMT
Sbjct: 20  MSIDDYRPTYMVEAIYDLRANDLLRHGITAVLVDLDNTLIAWNNPDGTPEVRAWLDEMTI 79

Query: 72  ADISVVVVSNNNHARVERAVSRFGVDFVSRAMKPFTRGINMAIERYGFDRDEVIMVGDQL 131
            ADISVVVVSNN H+RVERAVSRFGVDF+SRA+KPF  GI  AI RYGFDR+EVIMVGDQL
Sbjct: 80  ADISVVVVSNNKHSRVERAVSRFGVDFISRALKPFAYGIEKAIARYGFDRNEVIMVGDQL 139

Query: 132 MTDIRASHRAGIKSVLVKPIVKSDAWNTKFNRLRERRVWKKIEENYGKIVYQKGI 186
            MTDIRASHRAGIKSVLVKP+V SDAWNTK NR RERRV  K+EE YGK+ YQKGI
Sbjct: 140 MTDIRASHRAGIKSVLVKPLVASDAWNTKINRWRERRVMAKLEEKYGKLSYQKGI 194
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1799

A DNA sequence (GBSx1906) was identified in *S.agalactiae* <SEQ ID 5593> which encodes the amino acid sequence <SEQ ID 5594>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1091(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14509 GB:Z99117 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 220/373 (58%), Positives = 280/373 (74%), Gaps = 8/373 (2%)

Query: 1   MEELFCIGCGARIQTENKDAAGYTPRAALEKGLETGELYCQRCFRLRHYNEITDVHITDD 60
            ME++ CIGCG  IQTE+K   GY P A+L K       + CQRCFRL++YNEI DV +TDD
Sbjct: 1   MEKVVCIGCGVTIQTEDKTGLGYAPPASLTKE----NVICQRCFRLKNYNEIQDVSLTDD 56

Query: 61  EFLKLLHEVGDSDALVVNVIDIFDFNGSIIPGLSRFVAGNDVLLVGNKKDILPKSVKDGK 120
            +FL +LH +G++D+LVV ++DIFDFNGS I GL R V GN +LLVGNK DILPKS+K  +
Sbjct: 57  DFLNILHGIGETDSLVVKIVDIFDFNGSWINGLQRLVGGNPILLVGNKADILPKSLKRER 116

Query: 121 VTQWLTERAHEEGLRPVDVILTSAQNHHAIKDLIDTIEKYRHGQDVYVVGVTNVGKSTLI 180
            + QW+   A E GL+PVDV L SA     I+++ID IE YR+G+DVYVVG TNVGKST I
Sbjct: 117 LIQWMKREAKELGLKPVDVFLVSAGRGQGIREVIDAIEHYRNGKDVYVVGCTNVGKSTFI 176

Query: 181 NAIIREITGSRDVITTSRFPGTTLDKIEIPLDDGSYIFDTPGIIHRHQMAHYLTAKNLKY 240
            N II+E++G  D+ITTS+FPGTTLD IEIPLDDGS ++DTPGII+ HQMAHY+  K+LK
Sbjct: 177 NRIIKEVSGEEDIITTSQFPGTTLDAIEIPLDDGSSLYDTPGIINNHQMAHYVNKKDLKI 236

Query: 241 VSPKKEIKPKTYQLNSEQTLFLAGLARFDFISGQKQGFTAYFDNNLNLHRTKLVGADEFY 300
            +SPKKE+KP+T+QLN +QTL+  GLARFD++SG++  F  Y  N L +HRTKL  AD  Y
Sbjct: 237 LSPKKELKPRTFQLNDQQTLYFGGLARFDYVSGERSPFICYMPNELMIHRTKLENADALY 296

Query: 301 TKHVGKLLTPPTGKEVSDFPKLVRHEFTIKD-KMDIVYSGLGWIRVKSEAENPVVVAAWA 359
```

```
            KH G+LLTPP   E+ +FP+LV H FTIKD K DIV+SGLGW+ V    +    V A+A
Sbjct: 297 EKHAGELLTPPGKDEMDEFPELVAHTFTIKDKKTDIVFSGLGWVTVHDADKK---VTAYA 353

Query: 360 PEGVAVVLRKALI 372
            P+GV V +R++LI
Sbjct: 354 PKGVHVFVRRSLI 366
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5595> which encodes the amino acid sequence <SEQ ID 5596>. Analysis of this protein sequence reveals the following:

```
    Possible site: 15

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB14509 GB:Z99117 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 220/373 (58%), Positives = 286/373 (75%), Gaps = 8/373 (2%)

Query: 1   MEELFCIGCGIQIQTEDKEKAGFTPAAALKKGMETGELYCQRCFRLRHYNEITDVHITDD 60
            ME++ CIGCG+ IQTEDK   G+ P A+L K      + CQRCFRL++YNEI DV +TDD
Sbjct: 1   MEKVVCIGCGVTIQTEDKTGLGYAPPASLTKE----NVICQRCFRLKNYNEIQDVSLTDD 56

Query: 61  EFLRLLHEVGDSDALVVNVIDIFDFNGSIIPGLSRFISGNDVLLVGNKKDILPKSVKDGK 120
            +FL +LH +G++D+LVV ++DIFDFNGS I GL R + GN +LLVGNK DILPKS+K  +
Sbjct: 57  DFLNILHGIGETDSLVVKIVDIFDFNGSWINGLQRLVGGNPILLVGNKADILPKSLKRER 116

Query: 121 VTQWLTERAHEEGLRPLDVMLTSAQNKYAIKDLIGRINELRNGRDVYVVGVTNVGKSTLI 180
            + QW+   A E GL+P+DV L SA     I+++I  I   RNG+DVYVVG TNVGKST I
Sbjct: 117 LIQWMKREAKELGLKPVDVFLVSAGRGQGIREVIDAIEHYRNGKDVYVVGCTNVGKSTFI 176

Query: 181 NAIIQEITGNKDVITTSRFPGTTLDKIEIPLDDGTFIFDTPGIIHRHQMAHYLSPKELKI 240
            N II+E++G +D+ITTS+FPGTTLD IEIPLDDG+ ++DTPGII+ HQMAHY++ K+LKI
Sbjct: 177 NRIIKEVSGEEDIITTSQFPGTTLDAIEIPLDDGSSLYDTPGIINNHQMAHYVNKKDLKI 236

Query: 241 VSPKKEIKPKTYQLNPEQTLFLGGLARFDFINGERQGFTAFFDNQLELHRTKLAGADAFY 300
            +SPKKE+KP+T+QLN +QTL+ GGLARFD+++GER  F  +  N+L +HRTKL  ADA Y
Sbjct: 237 LSPKKELKPRTFQLNDQQTLYFGGLARFDYVSGERSPFICYMPNELMIHRTKLENADALY 296

Query: 301 DKHVGTLLTPPDKKELTAFPKLVRHEFTI-DQKMDIVFSGLGWIRVNGQKDSKAIVAAWA 359
            +KH G LLTPP K E+  FP+LV H FTI D+K DIVFSGLGW+ V+   D+   V A+A
Sbjct: 297 EKHAGELLTPPGKDEMDEFPELVAHTFTIKDKKTDIVFSGLGWVTVH---DADKKVTAYA 353

Query: 360 PEGVAVIVRKAII 372
            P+GV V VR+++I
Sbjct: 354 PKGVHVFVRRSLI 366
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 308/372 (82%), Positives = 343/372 (91%)

Query: 1   MEELFCIGCGARIQTENKDAAGYTPRAALEKGLETGELYCQRCFRLRHYNEITDVHITDD 60
            MEELFCIGCG +IQTE+K+ AG+TP AAL+KG+ETGELYCQRCFRLRHYNEITDVHITDD
Sbjct: 1   MEELFCIGCGIQIQTEDKEKAGFTPAAALKKGMETGELYCQRCFRLRHYNEITDVHITDD 60

Query: 61  EFLKLLHEVGDSDALVVNVIDIFDFNGSIIPGLSRFVAGNDVLLVGNKKDILPKSVKDGK 120
            EFL+LLHEVGDSDALVVNVIDIFDFNGSIIPGLSRF++GNDVLLVGNKKDILPKSVKDGK
Sbjct: 61  EFLRLLHEVGDSDALVVNVIDIFDFNGSIIPGLSRFISGNDVLLVGNKKDILPKSVKDGK 120

Query: 121 VTQWLTERAHEEGLRPVDVILTSAQNHHAIKDLIDTIEKYRHGQDVYVVGVTNVGKSTLI 180
            VTQWLTERAHEEGLRP+DV+LTSAQN +AIKDLI  I + R+G+DVYVVGVTNVGKSTLI
Sbjct: 121 VTQWLTERAHEEGLRPLDVMLTSAQNKYAIKDLIGRINELRNGRDVYVVGVTNVGKSTLI 180
```

```
Query: 181 NAIIREITGSRDVITTSRFPGTTLDKIEIPLDDGSYIFDTPGIIHRHQMAHYLTAKNLKY 240
           NAII+EITG++DVITTSRFPGTTLDKIEIPLDDG++IFDTPGIIHRHQMAHYL+ K LK
Sbjct: 181 NAIIQEITGNKDVITTSRFPGTTLDKIEIPLDDGTFIFDTPGIIHRHQMAHYLSPKELKI 240

Query: 241 VSPKKEIKPKTYQLNSEQTLFLAGLARFDFISGQKQGFTAYFDNNLNLHRTKLVGADEFY 300
           VSPKKEIKPKTYQLN EQTLFL GLARFDFI+G++QGFTA+FDN L LHRTKL GAD FY
Sbjct: 241 VSPKKEIKPKTYQLNPEQTLFLGGLARFDFINGERQGFTAFFDNQLELHRTKLAGADAFY 300

Query: 301 TKHVGKLLTPPTGKEVSDFPKLVRHEFTIKDKMDIVYSGLGWIRVKSEAENPVVVAAWAP 360
            KHVG LLTPP  KE++ FPKLVRHEFTI  KMDIV+SGLGWIRV  + ++  +VAAWAP
Sbjct: 301 DKHVGTLLTPPDKKELTAFPKLVRHEFTIDQKMDIVFSGLGWIRVNGQKDSKAIVAAWAP 360

Query: 361 EGVAVVLRKALI 372
           EGVAV++RKA+I
Sbjct: 361 EGVAVIVRKAII 372
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1800

A DNA sequence (GBSx1907) was identified in *S.agalactiae* <SEQ ID 5597> which encodes the amino acid sequence <SEQ ID 5598>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2948(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14507 GB:Z99117 similar to dihydrodipicolinate reductase
          [Bacillus subtilis]
 Identities = 49/97 (50%), Positives = 67/97 (68%), Gaps = 2/97 (2%)

Query: 1  MLTSKQRAFLKSEAHSMKPIIQIGKNGLNDQIKTSVRNALDARELIKVTLLQNTDEDIHD 60
          MLT KQ+ FL+S+AH + PI Q+GK G+ND +   +  AL+ARELIKV++LQN +ED +D
Sbjct: 1  MLTGKQKRFLRSKAHHLTPIFQVGKGGVNDNMIKQIAEALEARELIKVSVLQNCEEDKND 60

Query: 61 VAEVLEDEIGCDTVLKIGRILILYKESARKENRKISV 97
          VAE L        V  IG  ++LYKES  KEN++I +
Sbjct: 61 VAEALVKGSRSQLVQTIGNTIVLYKES--KENKQIEL 95
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5599> which encodes the amino acid sequence <SEQ ID 5600>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2839(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 89/102 (87%), Positives = 98/102 (95%)

Query: 1   MLTSKQRAFLKSEAHSMKPIIQIGKNGLNDQIKTSVRNALDARELIKVTLLQNTDEDIHD 60
           MLTSKQRAFLKSEAHS+KPI+QIGKNGLND IKTS+R ALDARELIKVTLLQNTDEDIH+
Sbjct: 1   MLTSKQRAFLKSEAHSLKPIVQIGKNGLNDHIKTSIRQALDARELIKVTLLQNTDEDIHE 60
```

```
Query: 61  VAEVLEDEIGCDTVLKIGRILILYKESARKENRKISVKVKAV 102
           VAE+LE+EIGCDTVLKIGRILILYK SA+KENRK+S KVKA+
Sbjct: 61  VAEILEEEIGCDTVLKIGRILILYKVSAKKENRKLSPKVKAI 102
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1801

A DNA sequence (GBSx1908) was identified in *S.agalactiae* <SEQ ID 5601> which encodes the amino acid sequence <SEQ ID 5602>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.66    Transmembrane     3 -  19 (    1 -  21)

----- Final Results -----
              bacterial membrane --- Certainty=0.2062(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10089> which encodes amino acid sequence <SEQ ID 10090> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14506 GB:Z99117 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 85/187 (45%), Positives = 134/187 (71%)

Query: 38  KQIGIMGGNFNPVHNAHLVVADQVRQQLCLDQVLLMPEFQPPHIDKKETIDEQHRLKMLE 97
           K+IGI GG F+P HN HL++A++V  Q  LD++  MP   PPH   ++  D  HR++ML+
Sbjct: 2   KKIGIFGGTFDPPHNGHLLMANEVLYQAGLDEIWFMPNQIPPHKQNEDYTDSFHRVEMLK 61

Query: 98  LAIEGIDGLSIEPIEIERKGISYTYDTMKLLIEKNPDVDYYFIIGADMVEYLPKWHRIDE 157
           LAI+      +E +E+ER+G SYT+DT+ LL ++ P+   +FIIGADM+EYLPKW+++DE
Sbjct: 62  LAIQSNPSFKLELVEMEREGPSYTFDTVSLLKQRYPNDQLFFIIGADMIEYLPKWYKLDE 121

Query: 158 LVKMVQFVGVQRPKYKAGTSYPVIWVDLPLMDISSSMIRQFIKSNRQPNYLLPREVLDYI 217
           L+ ++QF+GV+RP +   T YP+++ D+P  ++SS+MIR+  KS +  +YL+P +V  Y+
Sbjct: 122 LLNLIQFIGVKRPGFHVETPYPLLFADVPEFEVSSTMIRERFKSKKPTDYLIPDKVKKYV 181

Query: 218 RKEGLYK 224
            + GLY+
Sbjct: 182 EENGLYE 188
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5603> which encodes the amino acid sequence <SEQ ID 5604>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.4660(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 172/210 (81%), Positives = 196/210 (92%)

Query: 15  MALELLTPFTKVELEEKKRDTNRKQIGIMGGNFNPVHNAHLVVADQVRQQLCLDQVLLMP 74
           MALELLTPFTKVELEE+K+++NRKQIGI+GGNFNP+HNAHLVVADQVRQQL LDQVLLMP
Sbjct: 1   MALELLTPFTKVELEEEKKESNRKQIGILGGNFNPIHNAHLVVADQVRQQLGLDQVLLMP 60
```

```
Query: 75  EFQPPHIDKKETIDEQHRLKMLELAIEGIDGLSIEPIEIERKGISYTYDTMKLLIEKNPD 134
           E +PPH+D KETIDE+HRL+MLELAIE ++GL+IE  E+ER+GISYTYDTM  L E++PD
Sbjct: 61  ECKPPHVDAKETIDEKHRLRMLELAIEDVEGLAIETCELERQGISYTYDTMLYLTEQHPD 120

Query: 135 VDYYFIIGADMVEYLPKWHRIDELVKMVQFVGVQRPKYKAGTSYPVIWVDLPLMDISSSM 194
           VD+YFIIGADMV+YLPKWHRIDELVK+VQFVGVQRPKYKAGTSYPVIWVDLPL+DISSSM
Sbjct: 121 VDFYFIIGADMVDYLPKWHRIDELVKLVQFVGVQRPKYKAGTSYPVIWVDLPLIDISSSM 180

Query: 195 IRQFIKSNRQPNYLLPREVLDYIRKEGLYK 224
           IR FIK  RQPNYLLP+ VLDYI +EGLY+
Sbjct: 181 IRDFIKKGRQPNYLLPKRVLDYITQEGLYQ 210
```

SEQ ID 5602 (GBS651) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 132 (lane 8-10; MW 53.3kDa) and in Figure 186 (lane 8; MW 53kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 132 (lane 12; MW 28.4kDa) and in Figure 140 (lane 11; MW 20kDa).

Purified GBS651-GST is shown in Figure 243, lane 4; purified GBS651-His is shown in Fig.229, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1802

A DNA sequence (GBSx1909) was identified in *S.agalactiae* <SEQ ID 5605> which encodes the amino acid sequence <SEQ ID 5606>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4281(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14505 GB:Z99117 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 79/180 (43%), Positives = 115/180 (63%)

Query: 9   LDRTELLSKVRHMMSDKRFNHVLGVERAAIELAERYGYDKEKAGLAALLHDYAKELSDDE 68
           ++R E L+ V+  +++ R+ H +GV   AIELAER+G D +KA +AA+ HDYAK    +E
Sbjct: 1   MNREEALACVKQQLTEHRYIHTVGVMNTAIELAERFGADSKKAEIAAIFHDYAKFRPKEE 60

Query: 69  FLRLIDKYQPDPDLKKWGNNIWHGLVGIYKIQEDLAIKDQDILAAIAKHTVGSAQMSTLD 128
             ++I + +    L      +WH  VG Y +Q +  ++D+DIL AI  HT G   M+ L+
Sbjct: 61  MKQIIAREKMPAHLLDHNPELWHAPVGAYLVQREAGVQDEDILDAIRYHTSGRPGMTLLE 120

Query: 129 KIVYVADYIEHNRDFPGVEEARELAKVDLNKAVAYETARTVAFLASKAQPIYPKTIETYN 188
           K++YVADYIE NR FPGV+E R+LA+ DLN+A+      T+ FL  K QP++P T  TYN
Sbjct: 121 KVIYVADYIEPNRAFPGVDEVRKLAETDLNQALIQSIKNTMVFLMKKNQPVFPDTFLTYN 180
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5607> which encodes the amino acid sequence <SEQ ID 5608>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2615(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 130/194 (67%), Positives = 159/194 (81%)

Query: 1   MTYKDYTGLDRTELLSKVRHMMSDKRFNHVLGVERAAIELAERYGYDKEKAGLAALLHDY 60
           MTY+DY    RTELL+K+   MS KRF HVLGVE+AA+ LAE YG + +KAGLAALLHDY
Sbjct: 1   MTYEDYLPYSRTELLAKIAEQMSPKRFKHVLGVEKAALSLAECYGCNPDKAGLAALLHDY 60

Query: 61  AKELSDDEFLRLIDKYQPDPDLKKWGNNIWHGLVGIYKIQEDLAIKDQDILAAIAKHTVG 120
           AKE  D  FL LIDKYQ  P+L KW NN+WHG+VGIYKIQEDL +KD+DIL AI  HTVG
Sbjct: 61  AKECPDQVFLDLIDKYQLSPELAKWNNNVWHGMVGIYKIQEDLGLKDKDILRAIEIHTVG 120

Query: 121 SAQMSTLDKIVYVADYIEHNRDFPGVEEARELAKVDLNKAVAYETARTVAFLASKAQPIY 180
           +A+M+ LDK++YVADYIE  R FP V++AR++AK+DLN+AVAYET  TVA+LASKAQPI+
Sbjct: 121 AAEMTLLDKVLYVADYIEEGRIFPLVDDARKIAKLDLNQAVAYETVNTVAYLASKAQPIF 180

Query: 181 PKTIETYNAYIPYL 194
           P+T++TYNA+  YL
Sbjct: 181 PQTLDTYNAFCSYL 194
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1803

A DNA sequence (GBSx1910) was identified in *S.agalactiae* <SEQ ID 5609> which encodes the amino acid sequence <SEQ ID 5610>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -2.34    Transmembrane    12 -  28 (  10 -  28)

----- Final Results -----
            bacterial membrane --- Certainty=0.1935(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10087> which encodes amino acid sequence <SEQ ID 10088> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG19496 GB:AE005041 Vng1100c [Halobacterium sp. NRC-1]
 Identities = 46/175 (26%), Positives = 82/175 (46%), Gaps = 12/175 (6%)

Query: 22  ALLLIDIQQGIMDKK--PKHLTNFAVLLDDLLLSAKGSNCEVIWIRHHDKE----LPQGS 75
           AL+L+D QQG  D     ++  +     ++LL + + +   +  +RH+  E    L QG
Sbjct: 7   ALVLVDFQQGFADPAWGDRNNPDAEAHAEELLAAWRDAAAPIAHVRHNSTEATSPLRQGE 66

Query: 76  PQWEIWEQRHLVTHHKIIDKTYNSCFKDTHLHDYLQSKHISQLIMMGLQTEYCFDTSVKV 135
           P +   +            K+ N  F DT L  +L+ +    L++ GL T++C  T+V++
Sbjct: 67  PGFAYTDGLAPAADEPEFVKSVNGAFVDTALEGWLRDRDTGSLVVCGLTTDHCVSTTVRM 126

Query: 136 AFEYGYDIFIPQGGHLTFDTPTLSGDSIKK---HYENIWHHR--FATMVAKDSLL 185
           A   G+D+ + +    T D  TL G+ +     H   + H R  FAT+    ++L
Sbjct: 127 ADNRGFDVTLVRDATATHDR-TLDGERLPPSVVHRTALAHLRGEFATLATTATVL 180
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 5610 (GBS652) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 133 (lane 2 & 3; MW 49.7kDa) + lane 4; MW 27kDa) and in Figure 186 (lane 9; MW 50kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 133 (lane 5 & 7; MW 24.8kDa) and in Figure 178 (lane 10; MW 25kDa). Purified GBS652-GST is shown in Figure 243, lane 9; purified GBS652-His is shown in Figure 229, lane 10.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1804

A DNA sequence (GBSx1911) was identified in *S.agalactiae* <SEQ ID 5611> which encodes the amino acid sequence <SEQ ID 5612>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0945(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14504 GB:Z99117 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 55/118 (46%), Positives = 82/118 (68%)

Query: 1    MTEKDLLQLVVKAADEKRAEDIVILDLQPVTSVADYFVIMSASNSRQLEAIADNIREQVK 60
             M +K +L++   A D+KRAEDI+ LD++ ++ VADYF+I    ++ +Q++AIA  I++Q
Sbjct: 1    MNQKSILKIAAAACDDKRAEDILALDMEGISLVADYFLICHGNSDKQVQAIAREIKDQAD 60

Query: 61   GNGGDASHLEGDSKAGWVLLDLNSVVVHIFSEDERQHYNLEKLWHEAPLLDAEVFMTE 118
              NG     +EG  +A WVL+DL  VVVH+F +DER +YNLEKLW +APL D +  M +
Sbjct: 61   ENGIQVKKMEGFDEARWVLVDLGDVVVHVFHKDERSYYNLEKLWGDAPLADLDFGMNQ 118
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5613> which encodes the amino acid sequence <SEQ ID 5614>. Analysis of this protein sequence reveals the following:

```
     Possible site: 50
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.69    Transmembrane   91 - 107 (  91 - 107)

----- Final Results -----
               bacterial membrane --- Certainty=0.1277(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB14504 GB:Z99117 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 55/113 (48%), Positives = 80/113 (70%)

Query: 17   MKKEELLKIVVEATEEKRAKDILALDLEGLTSLTDYFVIASATNSRQLEAIADNIREKVK 76
             M ++ +LKI   A ++KRA+DILALD+EG++ + DYF+I    + +Q++AIA  I+++
Sbjct: 1    MNQKSILKIAAAACDDKRAEDILALDMEGISLVADYFLICHGNSDKQVQAIAREIKDQAD 60

Query: 77   EAGGDASHVEGNSQAGWVLLDLTDVVVHLFLEDERYHYNLEKLWHEAPAVALD 129
             E G     +EG  +A WVL+DL DVVVH+F +DER +YNLEKLW +AP   LD
Sbjct: 61   ENGIQVKKMEGFDEARWVLVDLGDVVVHVFHKDERSYYNLEKLWGDAPLADLD 113
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 78/116 (67%), Positives = 100/116 (85%)

Query: 1    MTEKDLLQLVVKAADEKRAEDIVILDLQPVTSVADYFVIMSASNSRQLEAIADNIREQVK 60
```

```
           M +++LL++VV+A +EKRA+DI+ LDL+ +TS+ DYFVI SA+NSRQLEAIADNIRE+VK
Sbjct: 17  MKKEELLKIVVEATEEKRAKDILALDLEGLTSLTDYFVIASATNSRQLEAIADNIREKVK 76

Query: 61  GNGGDASHLEGDSKAGWVLLDLNSVVVHIFSEDERQHYNLEKLWHEAPLLDAEVFM 116
             GGDASH+EG+S+AGWVLLDL  VVVH+F EDER HYNLEKLWHEAP +  + ++
Sbjct: 77  EAGGDASHVEGNSQAGWVLLDLTDVVVHLFLEDERYHYNLEKLWHEAPAVALDAYL 132
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1805

A DNA sequence (GBSx1912) was identified in *S.agalactiae* <SEQ ID 5615> which encodes the amino acid sequence <SEQ ID 5616>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2415(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1806

A DNA sequence (GBSx1913) was identified in *S.agalactiae* <SEQ ID 5617> which encodes the amino acid sequence <SEQ ID 5618>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1570(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14503 GB:Z99117 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 86/242 (35%), Positives = 154/242 (63%), Gaps = 4/242 (1%)

Query: 4   YETFAAVYDAVMDDTLYAKWTDFSLRHFPKGKKKLLELACGTGIQSVRFAQAGYAVTGLD 63
           Y+ FA+VYD +M    Y +WT +     P+ K ++L+LACGTG  S+R A+ G+ VTG+D
Sbjct: 3   YQGFASVYDELMSHAPYDQWTKWIEASLPE-KGRILDLACGTGEISIRLAEKGFEVTGID 61

Query: 64  LSGDMLKLAKKRATSAHQSIQFIEGNMLDLSNV-GKYDLITCYSDSICYMQDEVEVGDVF 122
           LS +ML  A+++ +S+ Q I F++ +M +++   G++D +    DS+ Y++ + +V + F
Sbjct: 62  LSEEMLSFAQQKVSSS-QPILFLQQDMREITGFDGQFDAVVICCDSLNYLKTKNDVIETF 120

Query: 123 IEVYKALEENGVFIFDVHSTYQTDKVFPGYSYHENADDFAMVWDTYEDDAPHSIVHELTF 182
             V++ L+  G+ +FDVHS+++  +VFP  ++ +  +D + +W ++      S++H+++F
Sbjct: 121 KSVFRVLKPEGILLFDVHSSFKIAEVFPDSTFADQDEDISYIWQSFAGSDELSVIHDMSF 180

Query: 183 FVQEEDGRFTRHDEVHEERTYDILTYDILLEQAGFKDVKVYADFEDKKPTATSARWFFVA 242
           FV   +  + R DE HE+RT+ +  Y+ +L+  GF+  +V ADF D +P+A S R FF A
Sbjct: 181 FVWNGEA-YDRFDETHEQRTFPVEEYEEMLKNCGFQLHRVTADFTDTEPSAQSERLFFKA 239
```

```
Query: 243 HK 244
            K
Sbjct: 240 QK 241
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5619> which encodes the amino acid sequence <SEQ ID 5620>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2315(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 191/243 (78%), Positives = 215/243 (87%), Gaps = 2/243 (0%)

Query: 4   YETFAAVYDAVMDDTLYAKWTDFSLRHFPK--GKKKLLELACGTGIQSVRFAQAGYAVTG 61
           YE FA+VYDAVMDD+LY  WTDFSLRH PK  G+ +LLELACGTGIQSVRFAQAG+ VTG
Sbjct: 21  YEKFASVYDAVMDDSLYDLWTDFSLRHLPKSKGRNRLLELACGTGIQSVRFAQAGFDVTG 80

Query: 62  LDLSGDMLKLAKKRATSAHQSIQFIEGNMLDLSNVGKYDLITCYSDSICYMQDEVEVGDV 121
           LDLS DML +AKKRA SA + I FI+GNMLDLS VG++D +TCYSDSICYMQDEV+VGDV
Sbjct: 81  LDLSQDMLAIAKKRAQSAKKKIDFIQGNMLDLSQVGQFDFVTCYSDSICYMQDEVDVGDV 140

Query: 122 FIEVYKALEENGVFIFDVHSTYQTDKVFPGYSYHENADDFAMVWDTYEDDAPHSIVHELT 181
           F EVY  L  +G+FIFDVHSTYQTD+ FPGYSYHENADDFAMVWDTY D+APHS+VHELT
Sbjct: 141 FKEVYDVLANDGIFIFDVHSTYQTDECFPGYSYHENADDFAMVWDTYADEAPHSVVHELT 200

Query: 182 FFVQEEDGRFTRHDEVHEERTYDILTYDILLEQAGFKDVKVYADFEDKKPTATSARWFFV 241
           FF+QE+DGRF+R DEVHEERTY++LTYDILLEQAGFK  KVYADFEDK+PT TS RWFFV
Sbjct: 201 FFIQEDDGRFSRFDEVHEERTYELLTYDILLEQAGFKSFKVYADFEDKEPTKTSKRWFFV 260

Query: 242 AHK 244
           A+K
Sbjct: 261 AYK 263
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1807

A DNA sequence (GBSx1914) was identified in *S.agalactiae* <SEQ ID 5621> which encodes the amino acid sequence <SEQ ID 5622>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3538(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06304 GB:AP001516 unknown conserved protein [Bacillus halodurans]
 Identities = 129/367 (35%), Positives = 184/367 (49%), Gaps = 45/367 (12%)

Query: 1   MTVTGIVAEFNPFHNGHKYLLEQAQ-----GIKVIAMSGNFMQRGEPAIVDKWTRSQMAL 55
           M   G+V E+NPFHNGH + L +A+      + +  MSG F+QRGEPAI+ KW R+ +AL
Sbjct: 1   MKAVGVVVEYNPFHNGHLHHLTEARKQAKADVVIAVMSGYFLQRGEPAILPKWERTSLAL 60
```

```
Query: 56  ENGADLVIELPFLVSVQSADYFASGAVSILARLGVDNLCFGTEE--MLDYARIGDIYVNK 113
           + GADLV+ELP+  S Q A++FA+GAVSILA L  D LCFG+EE  +  + R+
Sbjct: 61  QGGADLVVELPYAFSTQKAEWFATGAVSILAALEADALCFGSEEGTIEPFHRLYHFMAKH 120

Query: 114 KEEMEAFLKKQSD-SLSYPQKMQAMWQEFAGIT--FSGQTPNHILGLAYTKAA--SQNGI 168
           +   +  +K++ D  +SYP      ++   G         PN+ILG  Y KA       I
Sbjct: 121 RLAWDRMIKEELDKGMSYPTATSLAFKRLEGSAEHLDLSRPNNILGFHYVKAIYDLHTSI 180

Query: 169 RLNPIQRQGAGYHSSEKTE-IFASATSLRK--------HQSDRFF------VEKGMPNSD 213
           +   I R  AGYH     E   ASATS+RK           DR         + K
Sbjct: 181 KAMTIPRIKAGYHDDSLNESSIASATSIRKSLKTKEGWQMVDRVVPSYTTEMLKSFEKET 240

Query: 214 LFLNSPQVVWQDYFSLLKYQIMTHS--DLTQIYQVNEEIANRIKSQIRYVETVDELVDKV 271
            FL S    W+  F LLKY+++T +   L  IY+  E +  R    I    +  + + K+
Sbjct: 241 TFLPS----WERLFPLLKYRLLTATPEQLHAIYEGEEGLEYRALKTIVSATSFHDWMTKM 296

Query: 272 ATKRYTKARIRRLLTYILINAVESPIPNA----------IHVLGFTQKGQQHLKSVKK-- 319
            TKRYT  RI+R  T++  N  +  I +           I +LG T +GQ +L   KK
Sbjct: 297 KTKRYTWTRIQRYATHLFTNTTKEEIHSVLPRGTESLPYIRLLGMTSRGQMYLNGKKKQL 356

Query: 320 SVDIVTR 326
           +  ++TR
Sbjct: 357 TTPVITR 363
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5623> which encodes the amino acid sequence <SEQ ID 5624>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3165(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 221/359 (61%), Positives = 288/359 (79%)

Query: 1   MTVTGIVAEFNPFHNGHKYLLEQAQGIKVIAMSGNFMQRGEPAIVDKWTRSQMALENGAD 60
           MTVTGI+AEFNPFHNGHKYLLE A+G+K+IAMSGNFMQRGEPA++DKW RS+MAL+NGAD
Sbjct: 1   MTVTGIIAEFNPFHNGHKYLLETAEGLKIIAMSGNFMQRGEPALIDKWIRSEMALKNGAD 60

Query: 61  LVIELPFLVSVQSADYFASGAVSILARLGVDNLCFGTEEMLDYARIGDIYVNKKEEMEAF 120
           +V+ELPF VSVQSADYFA GA+ IL +LG+  L FGTE ++DY ++  +Y  K E+M A+
Sbjct: 61  IVVELPFFVSVQSADYFAQGAIDILCQLGIQQLAFGTENVIDYQKLIKVYEKKSEQMTAY 120

Query: 121 LKKQSDSLSYPQKMQAMWQEFAGITFSGQTPNHILGLAYTKAASQNGIRLNPIQRQGAGY 180
           L    D+ SYPQK Q MW+ FAG+ FSGQTPNHILGL+Y KA++   I+L PI+RQGA Y
Sbjct: 121 LSTLEDTFSYPQKTQKMWEIFAGVKFSGQTPNHILGLSYAKASAGKHIQLCPIKRQGAAY 180

Query: 181 HSSEKTEIFASATSLRKHQSDRFFVEKGMPNSDLFLNSPQVVWQDYFSLLKYQIMTHSDL 240
           HS +K  + ASA+++R+H +D  F+   +PN+ L +N+P + W  YFS LKYQI+ HSDL
Sbjct: 181 HSKDKNHLLASASAIRQHLNDWDFISHSVPNAGLLINNPHMSWDHYFSFLKYQILNHSDL 240

Query: 241 TQIYQVNEEIANRIKSQIRYVETVDELVDKVATKRYTKARIRRLLTYILINAVESPIPNA 300
           T I+QVN+E+A+RIK  I+  + +D LVD VATKRYTKAR+RR+LTYIL+NA E  +P
Sbjct: 241 TSIFQVNDELASRIKKAIKVSQNIDHLVDTVATKRYTKARVRRILTYILVNAKEPTLPKG 300

Query: 301 IHVLGFTQKGQQHLKSVKKSVDIVTRIGSQTWDSLTQRADSVYQMGNANIAEQTWGRIP 359
           IH+LGFT KGQ HLK +KKS  ++TRIG++TWD +TQ+ADS+YQ+G+ +I EQ++GRIP
Sbjct: 301 IHILGFTSKGQAHLKKLKKSRPLITRIGAETWDEMTQKADSIYQLGHQDIPEQSFGRIP 359
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1808

A DNA sequence (GBSx1915) was identified in *S.agalactiae* <SEQ ID 5625> which encodes the amino acid sequence <SEQ ID 5626>. This protein is predicted to be transcriptional activator tipa. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3117(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15677 GB:Z99122 transcriptional regulator [Bacillus subtilis]
 Identities = 91/246 (36%), Positives = 144/246 (57%), Gaps = 14/246 (5%)

Query: 4   VKEISHISGISVRTLHYYDEIDLLSPSFVGENGYRYYDDESLIKLQEILLFKELEFPLKK 63
           VK+++ ISG+S+RTLH+YD I+LL+PS + + GYR Y D  L +LQ+IL FKE+ F L +
Sbjct: 5   VKQVAEISGVSIRTLHHYDNIELLNPSALTDAGYRLYSDADLERLQQILFFKEIGFRLDE 64

Query: 64  IKEIMDSPNYDRNQALLDQIRWLELKKQRLEEVIEHAK----SIQRGKNMSD---FTAYN 116
           IKE++D PN+DR  AL  Q   L  KKQR++E+I+       S+  G+ M+    F   +
Sbjct: 65  IKEMLDHPNFDRKAALQSQKEILMKKKQRMDEMIQTIDRTLLSVDGGETMNKRDLFAGLS 124

Query: 117 QEELEAFQ----EEARTRWGD--TDSYKEFENSHSKNDFSMISQAMSQIFKDFGQLKELS 170
            +++E  Q    +E R  +G    +  ++  +++S +D+  I      I++      +
Sbjct: 125 MKDIEEHQQTYADEVRKLYGKEIAEETEKRTSAYSADDWRTIMAEFDSIYRRIAARMKHG 184

Query: 171 PTDEKVQKQVQILQDYITAQFYNCTNDLLASLGIMYIQDERFQKSIDNWGGQGTALFVSK 230
           P D ++Q  V   +D+I    Y+CT D+   LG +YI DERF  SI+ + G+G A F+ +
Sbjct: 185 PDDAEIQAAVGAFRDHICQYHYDCTLDIFRGLGEVYITDERFTDSINQY-GEGLAAFLRE 243

Query: 231 AIDSYC 236
           AI  YC
Sbjct: 244 AIIIYC 249
```

There is also homology to SEQ ID 1712.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1809

A DNA sequence (GBSx1916) was identified in *S.agalactiae* <SEQ ID 5627> which encodes the amino acid sequence <SEQ ID 5628>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2590(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14597 GB:Z99117 yrkC [Bacillus subtilis]
 Identities = 56/129 (43%), Positives = 74/129 (56%), Gaps = 7/129 (5%)

Query: 2   KGFHGNIEKLTLGNTNFRQVLYTAEHCQLVLMTLPVGGEIGSEIHAENDQFFRFEAGHGK 61
           K F  NI + T  N  FR  L+T +H Q+ LM+L +G +IG EIH   DQF R E G G
```

```
Sbjct: 59  KPFVVNINRATKQNNTFRTALWTGKHFQVTLMSLGIGEDIGLEIHPNVDQFLRIEQGRGI 118

Query: 62  VVIDGN------EYEVADGDAIIVPAGAEHNVINTSETEMLKLYTIYSPAHHKDGIIRAT 115
            V +  +      +  V D  AI+VPAG  HNVINT  T  LKLY+IY+P +H  G +  T
Sbjct: 119 VKMGKSKDHLNFQRNVYDDSAIVVPAGTWHNVINTGNTP-LKLYSIYAPPNHPFGTVHET 177

Query: 116 REEAEENEE 124
            + +A   E+
Sbjct: 178 KADAVAAED 186
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1810

A DNA sequence (GBSx1917) was identified in *S.agalactiae* <SEQ ID 5629> which encodes the amino acid sequence <SEQ ID 5630>. This protein is predicted to be glycerol uptake facilitator (glpF). Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.08   Transmembrane  156 - 172 ( 153 - 180)
   INTEGRAL    Likelihood = -6.21   Transmembrane  135 - 151 ( 132 - 155)
   INTEGRAL    Likelihood = -4.09   Transmembrane   86 - 102 (  80 - 103)
   INTEGRAL    Likelihood = -3.93   Transmembrane  213 - 229 ( 212 - 230)
   INTEGRAL    Likelihood = -3.72   Transmembrane    8 -  24 (   5 -  28)
   INTEGRAL    Likelihood = -2.76   Transmembrane   38 -  54 (  36 -  58)

----- Final Results -----
               bacterial membrane --- Certainty=0.4630(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04811 GB:AP001510 glycerol uptake facilitator [Bacillus halodurans]
 Identities = 135/230 (58%), Positives = 171/230 (73%)

Query: 1   MTQFLGEFLGTFILVLLGDGVVAGNVLSKTKEEGTGWTAIVFGWGIACTVAVYVSGLFSP 60
            M+ FLGE +GT IL++LG GVVAG VL  TK E  GW  I   WG+A   AVY  G  S
Sbjct: 1   MSPFLGEVIGTMILIILGGGVVAGVVLKGTKSENGGWIVITAAWGLAVATAVYCVGQISG 60

Query: 61  AHLNPAVTLAMASIGAISWGQVIPFIIAQMLGAMVAATILWLHYYPHWKETKDSGLILAS 120
            AHLNPAVT+ +A +GA  W QV  +I+AQMLGAM+ AT+++LHYYPH+K T+D G  LA
Sbjct: 61  AHLNPAVTIGLALVGAFEWSQVAGYIVAQMLGAMIGATLVFLHYYPHFKATEDQGAKLAV 120

Query: 121 FSTGPAIRHTPSNLLGEIIGTAILVITIMAIGPSKVAAGLGPIIVGIVIFAVGFSLDPTT 180
            FST PAI+H P+N   E++GT +LV+ I+AIG ++   GL P+IVG++I  +G SL  TT
Sbjct: 121 FSTDPAIKHLPANFFSEVLGTFVLVLGILAIGANEFTEGLNPLIVGLLIVVIGLSLGGTT 180

Query: 181 GYAINPARDLGPRLMHAILPIENKGNSDWSYAWIPVVGPIIGGVLGAILY 230
            GYAINPARDLGPR+ H +LPI  KG+S+WSYAWIP+VGPIIGG +GA+ Y
Sbjct: 181 GYAINPARDLGPRIAHFLLPIPGKGSSNWSYAWIPIVGPIIGGGIGALTY 230
```

There is also homology to SEQ ID 2854.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1811

A DNA sequence (GBSx1918) was identified in *S.agalactiae* <SEQ ID 5631> which encodes the amino acid sequence <SEQ ID 5632>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1694(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07114 GB:AP001518 unknown conserved protein in others
          [Bacillus halodurans]
 Identities = 64/118 (54%), Positives = 85/118 (71%)

Query: 5   GIIVVSHSKNIAQGVVDLISEVAKDVSITYVGGTEDGEIGTSFDQVQQIVEQNDKKTLLA 64
           GI++ SH   +A+G+V L+ E AKDVSITY GGT+D ++G SF+++QQ V  N+   L
Sbjct: 7   GIVISSHVPALAEGIVTLLKEAAKDVSITYAGGTDDDQVGASFEKIQQAVMDNEADELFV 66

Query: 65  FFDLGSAKMNLELVADFSEKNIIINSVPVVEGAYTAAALLQAGADLDSIQSQLAELTI 122
           F+DLGSAKMN+E+V + SEK I +  V +VEGAYTAAAL Q GA  ++I  QL  LTI
Sbjct: 67  FYDLGSAKMNVEMVMELSEKTIHLMDVALVEGAYTAAALTQGGASFETIMEQLQPLTI 124
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1812

A DNA sequence (GBSx1919) was identified in *S.agalactiae* <SEQ ID 5633> which encodes the amino acid sequence <SEQ ID 5634>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4753(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07115 GB:AP001518 unknown conserved protein in others
          [Bacillus halodurans]
 Identities = 98/190 (51%), Positives = 135/190 (70%), Gaps = 2/190 (1%)

Query: 3   VKTAIEWMHTFNQKIQSNKDYLSELDTPIGDGDHGGNMARGMTAVIENLDNNEFSSAADV 62
           V+   +W+H F++K+Q+N+ YLSELD+ IGDGDHG NMARG+  V   L  N F S  +V
Sbjct: 4   VENTTKWLHAFHEKVQANQSYLSELDSAIGDGDHGTNMARGLAEVERKLKENLFESPQEV 63

Query: 63  FKTVSMQLLSKVGGASGPLYGSAFMGITK-AEQSKSTISEALGAGLEMIQKRGKAELNEK 121
            K  +M L+SK GGASGPLYG+A + ++K        I +++ AGL  I KRGKA   EK
Sbjct: 64  LKMAAMALISKTGGASGPLYGTALLEMSKQVANDPQNIGKSIEAGLNGILKRGKATTGEK 123

Query: 122 TMVDVWHGVIEAI-EKNELTEDRIDSLVDATKGMKATKGRASYVGERSVGHIDPGSFSSG 180
           TMVD+W  V+E++  + +L+++RI   V  TK MKATKGRASY+GERS+GH+DPG+ SSG
Sbjct: 124 TMVDIWKPVVESLMAEQQLSKERIQQFVSETKEMKATKGRASYLGERSLGHLDPGAVSSG 183

Query: 181 LLFKALLEVG 190
            LF+A+++ G
```

```
Sbjct: 184 YLFEAMIDGG 193
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1813

A DNA sequence (GBSx1920) was identified in *S.agalactiae* <SEQ ID 5637> which encodes the amino acid sequence <SEQ ID 5638>. This protein is predicted to be dihydroxyacetone kinase (b1200). Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2080(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07116 GB:AP001518 dihydroxyacetone kinase [Bacillus halodurans]
 Identities = 204/329 (62%), Positives = 261/329 (79%)

Query: 1   MKKILNQPTDVVTEMLDGLAYVHNDLVHRIEGFDIIARNEEKSGKVALISGGGSGHEPSH 60
            MKKILN P +V+ EMLDG  Y +  LV R+ G  +I R  E  GKVAL+SGGGSGHEPSH
Sbjct: 1   MKKILNDPQNVLDEMLDGFVYANGHLVERVAGTGVIRRTYEDKGKVALVSGGGSGHEPSH 60

Query: 61  AGFVGEGMLSAAVCGAVFTSPTPDQVLEAIKEADEGAGVFMVIKNYSGDIMNFEMAQDMA 120
            AGFVG+GMLSAAVCG VFTSPTPDQ+ E IK AD+G GV ++IKNY+GD+MNFEMA +MA
Sbjct: 61  AGFVGQGMLSAAVCGEVFTSPTPDQIFEGIKAADQGGGVLLIIKNYTGDVMNFEMAGEMA 120

Query: 121 EMEGIEVASVVVDDDIAVEDSLYTQGKRGVAGTILVHKILGHAARHGKSLQEIKAIADEL 180
            E EGI V  ++V+DDIAVEDS +T G+RGVAGTI+VHKI+G AA  G SLQ +K + + +
Sbjct: 121 EAEGITVDHIIVNDDIAVEDSSFTAGRRGVAGTIIVHKIVGAAAEAGLSLQSLKVLGETV 180

Query: 181 VPNIHTVGLALSGATVPEVGKPGFVLAEDEIEFGIGIHGEPGYRKEKMQPSKALATELVD 240
            + N  T+G+++  ATVP VGKPGF L +DE+E+G+GIHGEPGYRKEK++ SK +A EL+
Sbjct: 181 IENTKTIGVSILPATVPAVGKPGFELGDDEMEYGVGIHGEPGYRKEKLKSSKEIAEELIL 240

Query: 241 KLIESFDAKSGEKYGVLINGMGATPLMEQYVFANDVAKLLEDKGIEVNYKKLGNYMTSID 300
            KL E+F    G+KYGVL+NG+GATPLMEQYVF NDVA  L ++G+ + +KK+G++MTSID
Sbjct: 241 KLKEAFGWSKGDKYGVLVNGLGATPLMEQYVFMNDVANKLTEEGLNIQFKKVGSFMTSID 300

Query: 301 MAGLSLTLIKLENQEWLEALNSDVTTIAW 329
            MAG+SLTLIK+  ++WL+  N +V T+ W
Sbjct: 301 MAGVSLTLIKIVEEKWLDYWNHEVKTVDW 329
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1814

A DNA sequence (GBSx1921) was identified in *S.agalactiae* <SEQ ID 5639> which encodes the amino acid sequence <SEQ ID 5640>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1997(Affirmative) < succ>
```

```
         bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07113 GB:AP001518 unknown [Bacillus halodurans]
 Identities = 59/142 (41%), Positives = 82/142 (57%), Gaps = 5/142 (3%)

Query: 1   MTSSLITKKKIAKSFKRLFISQAFDKISVSDIMEDAGIRRQTFYNHFVDKYALLEWIFQT 60
            MT+S+ITKK IAK+FK L   Q F KISVSDIM  A +RRQTFY HF DK+ LL WI++
Sbjct: 1   MTNSIITKKVIAKAFKDLMEVQPFSKISVSDIMNRANMRRQTFYYHFQDKFELLHWIYKQ 60

Query: 61  ELSEQVTDNLDYISGFQLLSELLTFFKMNQEFYIKLFQIEDQNDFSSYFESYCEQLVDKL 120
            E  E   D L Y     +   L+ +F  NQ FY +   +  QN F+ Y   + + L
Sbjct: 61  ETKEHSIDFLAYDDIHTIFRHLMHYFYENQTFYQRAMVVNGQNGFTDYLYEHIQTL---Y 117

Query: 121 LSDYSKSNFNQKERVTFINYHS 142
            L++  +   +QK+R    +++S
Sbjct: 118 LNEIDRR--SQKDREFISSFYS 137
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5641> which encodes the amino acid sequence <SEQ ID 5642>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
         bacterial cytoplasm --- Certainty=0.2101(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 31/115 (26%), Positives = 58/115 (49%), Gaps = 6/115 (5%)

Query: 7   TKKKIAKSFKRLFISQAFDKISVSDIMEDAGIRRQTFYNHFVDKYALLEWIFQTELSEQV 66
            TK  +  +   L   Q+F+ ++VSD+ + AGI R TFY H+ DK+ ++   F+ +  + +
Sbjct: 8   TKAYVKTALTTLLTEQSFETLTVSDLTKKAGINRGTFYLHYTDKFDMMNH-FKNDTLDDL 66

Query: 67  TDNLD----YISGFQLLSELLTFFKMNQEFYIKLFQIEDQNDFSSYFESYCEQLV 117
               L+    Y    Q+L++ L++   ++EF   L  I     F    + +C Q +
Sbjct: 67  YRLLNQAEIYTDTRQVLNQTLSYLIEHREFITALATI-SYLKFPQLIKDFCYQFL 120
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1815

A DNA sequence (GBSx1922) was identified in *S.agalactiae* <SEQ ID 5643> which encodes the amino acid sequence <SEQ ID 5644>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
         bacterial cytoplasm --- Certainty=0.1974(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1816

A DNA sequence (GBSx1923) was identified in *S.agalactiae* <SEQ ID 5645> which encodes the amino acid sequence <SEQ ID 5646>. This protein is predicted to be dihydroxyacetone kinase (b1200). Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1806(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07112 GB:AP001518 dihydroxyacetone kinase [Bacillus halodurans]
 Identities = 141/285 (49%), Positives = 197/285 (68%), Gaps = 1/285 (0%)

Query: 45  IPILSGGGSGHEPAHFGYVGEGMLSAAISGPIFVPPCASDILETIRFINRGKGVFVIIKN 104
            +PI+SGGGSGHEP H GYVGEGML+AA+ G +FVPP A  +L  IR +++GKGV +IIKN
Sbjct: 46  VPIISGGGSGHEPGHLGYVGEGMLAAAVHGDVFVPPSAQQVLAAIRQMDQGKGVLLIIKN 105

Query: 105 FEADLEEFSQAIEQARQEGIPIKYIVSHDDISVET-SNFKIRHRGVAGTVLLHKIIGQAA 163
            F ADL  F  A  QAR EG  + +++ +DD+SVE+ ++F+ R RGVAG VL+HKIIG AA
Sbjct: 106 FVADLATFLSAEVQARAEGRDVAHVIVNDDVSVESDASFEKRRRGVAGAVLVHKIIGAAA 165

Query: 164 LEGASLDELEQLGLSLTTSMATLGVASKSATILGQHQPVFDIEEGYISFGIGIHGEPGYR 223
             EG SL+ L+++G  +  ++ATLGVA   A +  + +P F +EEG + FG+GIHGE GYR
Sbjct: 166 KEGYSLEALQEIGEQVVKNLATLGVALTHADLPERREPQFLLEEGEVYFGVGIHGEQGYR 225

Query: 224 TMPFVSMEHLANELVNKLKMKLRWQDGEAFILLINNLGGSSKMEELLFTNAVMEFLALDD 283
                VS E LA ELVNKLK   RW   + + +LIN LGG+  +E+ +F N V   LA+++
Sbjct: 226 KEKLVSSELLAVELVNKLKSLYRWDKNDQYAVLINGLGGTPLIEQYVFANDVRRLLAIEN 285

Query: 284 LQLPFIKTGHLITSLDMAGLSVTLCRVKDSRWIDYLKHKTDARAW 328
            L + F+K G  +TSL+M G+S+T+ ++ D +W+ +L    D   W
Sbjct: 286 LHVSFVKVGTQLTSLNMKGISLTMLKICDEQWVKWLYAPVDVAHW 330
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1817

A DNA sequence (GBSx1924) was identified in *S.agalactiae* <SEQ ID 5647> which encodes the amino acid sequence <SEQ ID 5648>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3902(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10085> which encodes amino acid sequence <SEQ ID 10086> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC75047 GB:AE000290 orf, hypothetical protein [Escherichia coli K12]
 Identities = 182/237 (76%), Positives = 201/237 (84%)

Query: 20  MGRKWANIVAKKTAKDGANSKVYAKFGVEIYVAAKQGEPDPESNSALKFVLDRAKQAQVP 79
            MGRKWANIVAKKTAKDGA SK+YAKFGVEIY AAKQGEPDPE N++LKFV++RAKQAQVP
Sbjct: 1   MGRKWANIVAKKTAKDGATSKIYAKFGVEIYAAAKQGEPDPELNTSLKFVIERAKQAQVP 60

Query: 80  KHVIDKAIDKAKGNTDETFVEGRYEGFGPNGSMIIVDTLTSNVNRTAANVRTAYGKNGGN 139
            KHVIDKAIDKAKG  DETFV+GRYEGFGPNGSMII +TLTSNVNRT ANVRT + K GGN
Sbjct: 61  KHVIDKAIDKAKGGGDETFVQGRYEGFGPNGSMIIAETLTSNVNRTIANVRTIFNKKGGN 120

Query: 140 MGASGSVSYLFDKKGVIVFAGDDADTVFEQLLEADVDVDDVEAEEGTITVYTAPTDLHKG 199
            +GA+GSVSY+FD  GVIVF G D D +FE LLEA+VDV DV  EEG I +YT PTDLHKG
Sbjct: 121 IGAAGSVSYMFDNTGVIVFKGTDPDHIFEILLEAEVDVRDVTEEEGNIVIYTEPTDLHKG 180

Query: 200 IQALRDNGVEEFQVTELEMIPQSEVVLEGDDLETFEKLIDALESDDDVQKVYHNVAD 256
            I AL+  G+ EF  TELEMI QSEV L  +DLE FE L+DALE DDDVQKVYHNVA+
Sbjct: 181 IAALKAAGITEFSTTELEMIAQSEVELSPEDLEIFEGLVDALEDDDDVQKVYHNVAN 237
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5649> which encodes the amino acid sequence <SEQ ID 5650>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2926(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 233/238 (97%), Positives = 236/238 (98%)

Query: 20  MGRKWANIVAKKTAKDGANSKVYAKFGVEIYVAAKQGEPDPESNSALKFVLDRAKQAQVP 79
            MGRKWANIVAKKTAKDGA SKVYAKFGVEIYVAAKQGEPDPE N+ALKFV+DRAKQAQVP
Sbjct: 1   MGRKWANIVAKKTAKDGATSKVYAKFGVEIYVAAKQGEPDPELNTALKFVIDRAKQAQVP 60

Query: 80  KHVIDKAIDKAKGNTDETFVEGRYEGFGPNGSMIIVDTLTSNVNRTAANVRTAYGKNGGN 139
            KHVIDKAIDKAKGNTDETFVEGRYEGFGPNGSMIIVDTLTSNVNRTAANVRTAYGKNGGN
Sbjct: 61  KHVIDKAIDKAKGNTDETFVEGRYEGFGPNGSMIIVDTLTSNVNRTAANVRTAYGKNGGN 120

Query: 140 MGASGSVSYLFDKKGVIVFAGDDADTVFEQLLEADVDVDDVEAEEGTITVYTAPTDLHKG 199
            MGASGSVSYLFDKKGVIVFAGDDAD+VFEQLLEADVDVDDVEAEEGTITVYTAPTDLHKG
Sbjct: 121 MGASGSVSYLFDKKGVIVFAGDDADSVFEQLLEADVDVDDVEAEEGTITVYTAPTDLHKG 180

Query: 200 IQALRDNGVEEFQVTELEMIPQSEVVLEGDDLETFEKLIDALESDDDVQKVYHNVADF 257
            IQALRDNGVEEFQVTELEMIPQSEVVLEGDDLETFEKLIDALESDDDVQKVYHNVADF
Sbjct: 181 IQALRDNGVEEFQVTELEMIPQSEVVLEGDDLETFEKLIDALESDDDVQKVYHNVADF 238
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1818

A DNA sequence (GBSx1925) was identified in *S.agalactiae* <SEQ ID 5651> which encodes the amino acid sequence <SEQ ID 5652>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2507(Affirmative) < succ>
```

```
                  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1819

A DNA sequence (GBSx1926) was identified in *S.agalactiae* <SEQ ID 5653> which encodes the amino acid sequence <SEQ ID 5654>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1523(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA20826 GB:AL031541 hypothetical protein SCI35.37 [Streptomyces
           coelicolor A3(2)]
 Identities = 73/178 (41%), Positives = 101/178 (56%), Gaps = 2/178 (1%)

Query: 35  VKNAGGLPVILPISEAESAKAYVEMIDKLIISGGQNVLPSYYGEEKIIESDDYSLARDIF 94
           V+ AGGL  +LP    E A A V  +D ++I+GG +V P  YG E    +   + ARD +
Sbjct: 37  VQRAGGLAAMLPPDAPEHAAATVARVDGVVIAGGPDVEPVRYGAEPDPRTGPPARARDTW 96

Query: 95  EFALVEEALKQNKPIFAICRGMQLVNVALGGTLNQSIDNHYQEPYIGFAHYLNVEKGSFL 154
           E AL+E AL    P+  ICRGMQL+NVALGGTL Q I+ H +   +   H +    G+
Sbjct: 97  ELALIEAALAARVPLLGICRGMQLLNVALGGTLVQHIERHAEVVGVFGGHPVRPVPGTLY 156

Query: 155 EGFISGDFKINSLHRQSVKLLAEGLIVSARDPRDGTVEAYESRT-EQCIIGVQWHPEL 211
            G +  +  + + H Q+V  L  GL+ SA    DGTVEA E  +    ++GVQWHPE+
Sbjct: 157 AGAVPEETFVPTYHHQAVDRLGSGLVASAH-AADGTVEALEMPSGSGWVLGVQWHPEM 213
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5655> which encodes the amino acid sequence <SEQ ID 5656>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1210(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 111/230 (48%), Positives = 145/230 (62%), Gaps = 3/230 (1%)

Query: 2   LTKPIIGITGNEREMSDIPGYYYDSVSRHISEGVKNAGGLPVILPISEAESAKAYVEMID 61
           +TKPIIGIT N+R    +    +        + V  +GGLP++LPI +  +AK YV M+D
Sbjct: 1   MTKPIIGITANQRLNMALDNLPWSYAPTGFVQAVTQSGGLPLLLPIGDEAAAKTYVSMVD 60

Query: 62  KLIISGGQNVLPSYYGEEKIIESDDYSLARDIFEFALVEEALKQNKPIFAICRGMQLVNV 121
           K+I+ GGQNV P YY EEK    DD+S  RD FE A+++EA+   KPI  ICRG QL+NV
Sbjct: 61  KIILIGGQNVDPKYYQEEKAAFDDDFSPERDTFELAIIKEAITLKKPILGICRGTQLMNV 120
```

```
Query: 122 ALGGTLNQSIDNHYQE-PYIGFAHYLNVEKGSFLEGFISGDFKINSLHRQSVKLLAEGLI 180
           ALGG LNQ ID+H+QE P    +H + +E  S L         INS HRQS+K +A+ L
Sbjct: 121 ALGGNLNQHIDSHWQEAPSDFLSHEMIIEPDSILYPIYGHKTLINSFHRQSLKTVAKDLK 180

Query: 181 VSARDPRDGTVEAYESRTEQC-IIGVQWHPELMLH-QIENQTLFGYFVNE 228
           V ARDPRDGT+EA  S  +    +GVQWHPEL+   + E+  LF  FVN+
Sbjct: 181 VIARDPRDGTIEAVISTNDAIPFLGVQWHPELLQGVRDEDLQLFRLFVND 230
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1820

A DNA sequence (GBSx1927) was identified in *S.agalactiae* <SEQ ID 5657> which encodes the amino acid sequence <SEQ ID 5658>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5794(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1821

A DNA sequence (GBSx1928) was identified in *S.agalactiae* <SEQ ID 5659> which encodes the amino acid sequence <SEQ ID 5660>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0524(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8905> which encodes amino acid sequence <SEQ ID 8906> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: 22   Crend: 4
McG: Discrim Score:      8.37
GvH: Signal Score (-7.5): -0.64
     Possible site: 21
>>> May be a lipoprotein
ALOM program   count: 0 value:   6.74 threshold:  0.0
   PERIPHERAL  Likelihood =  6.74     112
 modified ALOM score:  -1.85

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear)
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2919> which encodes the amino acid sequence <SEQ ID 2920>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> May be a lipoprotein

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 120/162 (74%), Positives = 141/162 (86%), Gaps = 5/162 (3%)

Query: 6   LAACSSKSHTTKTGK----KEVNFATVGTTAPFSYVKDGKLTGFDIEVAKAVFKGSDNYK 61
           LAAC S S T ++G     KEV FATVGTTAPFSY K G+LTG+DIEVAKAVFKGSD+YK
Sbjct: 20  LAACGS-SKTAESGNQGSSKEVLFATVGTTAPFSYEKGGQLTGYDIEVAKAVFKGSDDYK 78

Query: 62  VTFKKTEWSSVFTGIDSGKFQMGGNNISYSSERSQKYLFSYPIGSTPSVLAVPKNSNIKA 121
           V+FKKTEWSS+FTG+DSGK+QMGGNNIS++ ERS KYLFSYPIGSTPSVL VPK+S+IK+
Sbjct: 79  VSFKKTEWSSIFTGLDSGKYQMGGNNISFTKERSAKYLFSYPIGSTPSVLVVPKDSDIKS 138

Query: 122 YNDISGHKTQVVQGTTTAKQLENFNKEHQKNPVTLKYTNENL 163
           ++DI GH TQVVQGTT+  QLE+FNK+H  NPVTLK+TNEN+
Sbjct: 139 FDDIQGHTTQVVQGTTSVAQLEDFNKKHSDNPVTLKFTNENI 180
```

SEQ ID 8906 (GBS71) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 4; MW 31.8kDa).

GBS71-His was purified as shown in Figure 196, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1822

A DNA sequence (GBSx1929) was identified in *S.agalactiae* <SEQ ID 5661> which encodes the amino acid sequence <SEQ ID 5662>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2179(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

There is also homology to SEQ ID 2920:

```
Identities = 64/91 (70%), Positives = 78/91 (85%)

Query: 1   MSDGKADFKLFDGPTVNAIIKNQGLTNLKTIPLTMRDQPYIYFIFGQDQKDLQKYVNNRL 60
           +S+GKADFK+FD PTVNAIIKNQGL NLKTI LT  +QP+IYFIF QDQ+ LQ +VN R+
Sbjct: 187 LSEGKADFKIFDAPTVNAIIKNQGLDNLKTIELTSTEQPFIYFIFSQDQEKLQSFVNKRI 246

Query: 61  KQLRKDGTLSKIAKEYLGGDYVPNEKDLVTP 91
           K+L  DGTLSK+AKE+LGGDYVP++K+L  P
Sbjct: 247 KELTADGTLSKLAKEHLGGDYVPSDKELKLP 277
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1823

A DNA sequence (GBSx1930) was identified in *S.agalactiae* <SEQ ID 5663> which encodes the amino acid sequence <SEQ ID 5664>. This protein is predicted to be 28 kDa outer membrane protein (yaeC). Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.44    Transmembrane   25 -  41 (  25 -  42)

----- Final Results -----
               bacterial membrane --- Certainty=0.1574(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB59825 GB:AJ012388 hypothetical protein [Lactococcus lactis]
 Identities = 110/283 (38%), Positives = 175/283 (60%), Gaps = 13/283 (4%)

Query: 22  KLKHIVLGLALTTLLGV----TFSNQEVSASSTSSKVVKVGVMTFSDTEKARWDKIEKLV 77
            K ++I++ +A+  L+ +    + ++Q    +S   K VKVG+M+    ++  W  +
Sbjct: 4   KNRNIIIAVAVLILVALVAFFSLNHQGGVKASAGEKTVKVGIMSGDKQDQEVWKSVANTA 63

Query: 78  GDK--AKIKFTEFTDYTQPNQATANKDVDINAFQHYNFLENWNKENKKNLIPLEKTYLAP 135
             +K   K+KF  F+DY QPN+A  + D+DINAFQ YN+++ WNK +K +++ +  TY+ P
Sbjct: 64  KEKYDLKLKFVYFSDYNQPNEALLSGDIDINAFQSYNYVKTWNKAHKSDIVAVGNTYITP 123

Query: 136 IRIYSEKVKSLKKLKKGATIAIPNDATNGSRALYVLQSAGLIKLNVS-GKKVATVANITS 194
            + IYS+++  L  LK+G+T+AIPNDA+N SRAL+VLQSAGL+KL  S   K+  + +IT
Sbjct: 124 MHIYSKEISKLSDLKEGSTVAIPNDASNESRALFVLQSAGLLKLTTSDSSKLVGLPDITE 183

Query: 195 NKKDINIQELDASQTPRALKDVDAAIINNTYIEQANLKPSDAIFVEKSDKNSKQWINIIA 254
            N   +  +E+DASQTPRAL  V  +++N  Y   A+L  S+++F+E  +K S Q+IN IA
Sbjct: 184 NPHQLKFKEVDASQTPRALDSVALSVVNYNYATAASLPKSESVFMEPLNKTSAQYINFIA 243

Query: 255 GRKNWKKQKNAKAIQAILDAYHTDEVKKVIKDTSAD---IPQW 294
                  K+KN K  + +  AY +   +K IK+   D   +P W
Sbjct: 244 ---TTSKEKNNKVYKEVAKAYASKATEKAIKEQYPDGGELPAW 283
```

There is also homology to SEQ ID 2132.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8907> and protein <SEQ ID 8908> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 4
McG: Discrim Score:      7.47
GvH: Signal Score (-7.5): -4.79
     Possible site: 21
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 1 value:  -1.44 threshold:  0.0
   INTEGRAL    Likelihood = -1.44    Transmembrane    5 -  21 (   5 -  22)
   PERIPHERAL  Likelihood =  5.20     147
 modified ALOM score:   0.79

*** Reasoning Step: 3

----- Final Results -----
```

```
   bacterial membrane --- Certainty=0.1574(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
40.6/63.1% over 279aa
                                                      Lactococcus lactis
  GP|6165402| hypothetical protein Insert characterized

ORF00442(364 - 1182 of 1482)
GP|6165402|emb|CAB59825.1||AJ012388(4 - 283 of 287) hypothetical protein {Lactococcus
lactis}
%Match = 21.0
%Identity = 40.6  %Similarity = 63.0
Matches = 112  Mismatches = 96  Conservative Sub.s = 62

162       192       222       252       282       312       342       372
WDTFKNS*RIPWR*LRTK*ERSRYS*GEVVIKTKEMSILSFLLYSLKL*QETVYNNLILITSYGIISLSQKLREFIMKLK
                                                                             | :
                                                                          MNPKNR

402                 450       480       510       540       564       594
HIVLGLALTTLLG--VTFS--NQEVSASSTSSKVVKVGVMTFSDTEKARWDKIEKLVGDK--AKIKFTEFTDYTQPNQAT
:|:: :|:  |:     ||  :|    :|   | ||||:|:    ::  |  :     :|   |:||  |:|| |||:|
NIIIAVAVLILVALVAFFSLNHQGGVKASAGEKTVKVGIMSGDKQDQEVWKSVANTAKEKYDLKLKFVYFSDYNQPNEAL
            20        30        40        50        60        70        80

624       654       684       714       744       774       804       834
ANKDVDINAFQHYNFLENWNKENKKNLIPLEKTYLAPIRIYSEKVKSLKKLKKGATIAIPNDATNGSRALYVLQSAGLIK
 : |:|||||| ||::: ||| :| ::: :  ||: |: |||:::  |  ||:|:|:||||||:| ||||:|||||||:|
LSGDIDINAFQSYNYVKTWNKAHKSDIVAVGNTYITPMHIYSKEISKLSDLKEGSTVAIPNDASNESRALFVLQSAGLLK
           100       110       120       130       140       150       160

861       891       921       951       981       1011      1041      1071
LNVS-GKKVATVANITSNKKDINIQELDASQTPRALKDVDAAIINNTYIEQANLKPSDAIFVEKSDKNSKQWINIIAGRK
|  |   |:  : :|| |   :  :|:|||||||||  |  :::|  |   |:|  |:::|:|  :| | |:|| ||
LTTSDSSKLVGLPDITENPHQLKFKEVDASQTPRALDSVALSVVNYNYATAASLPKSESVFMEPLNKTSAQYINFIA---
           180       190       200       210       220       230       240

1101      1131      1161      1182      1212      1242      1272      1302
NWKKQKNAKAIQAILDAYHTDEVKKVIKDTSAD---IPQW*RELTV*V*QGILIGYNLSAI*P*RAWDEYNVPGSWIVFE
   |:|| |  : :  || :   :| ||:   |   :| |  :|
TTSKEKNNKVYKEVAKAYASKATEKAIKEQYPDGGELPAWDLKL
              260       270       280
```

SEQ ID 8908 (GBS35) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 11 (lane 2; MW 31.6kDa).

The GBS35-His fusion product was purified (Figure 96A; see also Figure 192, lane 6) and used to immunise mice (lane 2 product; 20μg/mouse). The resulting antiserum was used for Western blot (Figure 96B), FACS (Figure 96C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Example 1824

A DNA sequence (GBSx1931) was identified in *S.agalactiae* <SEQ ID 5665> which encodes the amino acid sequence <SEQ ID 5666>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3126(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF11560 GB:AE002038 ArgE/DapE/Acy1 family protein [Deinococcus radiodurans]
 Identities = 129/419 (30%), Positives = 210/419 (49%), Gaps = 14/419 (3%)

Query: 26  LRDLIAIKSIFAQKVGLNDLSSYLGEVFIKAGAEVIIDDSYSAPFIVANFKSSKVDAKRI 85
            LR L+A+ S+ AQ   L + +  +  +    G  V       AP ++A           +
Sbjct: 16  LRALVALPSVSAQGRMLPETADAVAGLLRAEGFGVQQFPGTVAPVLLAEAGEGPFT---L 72

Query: 86  IFYNHYDTVPADEVEQWTEDPFTLSLRYGKMYGRGVDDDKGHITARLSAVKKYLSRHKGE 145
            + YNHYD  P D +E W   PF L+ R G++YGRG  DDKG + +RL+AV+  +    G
Sbjct: 73  LIYNHYDVQPEDPLELWDTPPFELTERGGRLYGRGASDDKGELASRLAAVRA-VREQLGH 131

Query: 146 LPLDITFIVEGAEESASVGLDYYLEKYQEQLQGADLIVWEDGPKNPKGQLEIAGGNKGIV 205
            LP+ I +++EG EE  S  L+ ++ ++  +LQ AD   WE G  +P+G+  ++ G KG++
Sbjct: 132 LPVKIKWLIEGEEEVGSPTLERFVAEHAAELQ-ADGCWWEFGGISPEGRPILSLGLKGVM 190

Query: 206 TFDLSVSSADVDIHSSFGGVVDSSTWYLIQALNTLRDNKGHILVEGIYDKVIPPTKRELE 265
              +L    AD D+HSS G V+D+  + L +A+ +LRD +G++ + G YD V   +  + +
Sbjct: 191 CLELRCRVADSDLHSSLGAVIDNPLYCLARAVASLRDEQGNVTIPGFYDDVRAASGADRQ 250

Query: 266 LVEKYSYRSAKALEGAYQLVLPSLADSHKTFLRKLYFEPSIAIEGITSGYQGEGVKTILP 325
             + +      +A+   + +  P    +   +  +    P + + G   GYQGEG KT+LP
Sbjct: 251 AIAQIP-GDGQAVRDTFGVRRP--LATGPAYNERTNLHPVVNVNGWGGGYQGEGSKTVLP 307

Query: 326 AYAKCKAEVRLVPGLTPKGVLDSIQNHLKENGFKDIELT-YTLGEMSYRSDMSAPSILKV 384
                 K + RLVP   P  VL  ++ HL   G  DIE+      +   R+D   P +
Sbjct: 308 GAGFVKLDFRLVPDQDPARVLSLLREHLTAQGLSDIEVVELEAHQKPARADAGHPFVQAC 367

Query: 385 VDLAEQFYPEGISLLPTSPGTGPMY-----LVHQALRAPIAAIGIGHANSRDHGVDENV 438
            V  A   + +   + P+S  +GPM+          L  P  A+GIG+   R H  +EN+
Sbjct: 368 VAAARAAHGQDPIVHPSSGASGPMFPFTGGAGGGGLGIPCVAVGIGNHAGRVHAPNENI 426
```

There is also homology to SEQ ID 2588.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1825

A DNA sequence (GBSx1932) was identified in *S.agalactiae* <SEQ ID 5667> which encodes the amino acid sequence <SEQ ID 5668>. This protein is predicted to be amino acid ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5366(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB59828 GB:AJ012388 hypothetical protein [Lactococcus lactis]
 Identities = 187/338 (55%), Positives = 256/338 (75%), Gaps = 12/338 (3%)

Query: 6   IIKLDNIDVTFHQKKREINAVKDVTIHINQGDIYGIVGYSGAGKSTLVRVINLLQEPSAG 65
            II+L+N+ V FHQK R + AVK+ T+HI +GDIYG++GYSGAGKSTLVR INLLQ+P+ G
Sbjct: 4   IIELNNLSVQFHQKGRLVTAVKNATLHIEKGDIYGVIGYSGAGKSTLVRTINLLQKPTEG 63

Query: 66  KITIDDQVIYD--NKVTLTSTQLREQRREIGMIFQHFNLMSQLTAEQNVAFALKHSG--- 120
            +I I+ + I+D  N V  T  +LRE R++IGMIFQHFNL+S+ T   NVAFAL+HS
```

```
Sbjct: 64  QIVINGEKIFDSENPVKFTGAKLREFRQKIGMIFQHFNLLSEKTVFNNVAFALQHSQIED 123

Query: 121 -------LSKEAKAAKVAKLLELVGLSDRAQNYPSQLSGGQKQRVAIARALANDPKILIS 173
                  L+K+ K  KV +LL+LV L+D +  YP+QLSGGQKQRVAIARALANDP+ILIS
Sbjct: 124 KNGKKRYLTKKEKNDKVTELLKLVDLADLSDKYPAQLSGGQKQRVAIARALANDPEILIS 183

Query: 174 DESTSALDPKTTKQILALLQDLNKKLGLTIVLITHEMQIVKDIANRVAVMQNGKLIEEGS 233
           DE TSALDPKTT QIL LL+ L++KLG+T+VLITHEMQ+VK+IAN+VAVMQNG++IE+ S
Sbjct: 184 DEGTSALDPKTTNQILDLLKSLHEKLGITVVLITHEMQVVKEIANKVAVMQNGEIIEQNS 243

Query: 234 VLDIFSHPRESLTQDFIKIATGIDEAMLKIEQQEVVKNLPVGSKLVQLKYAGHSTDEPLL 293
           ++DIF+ P+E+LT+ FI+  + ++  +  + + E++  L    +L+ L Y+G   ++P++
Sbjct: 244 LIDIFAQPKEALTKQFIETTSSVNRFIASLSKTELLAQLADDEELIHLDYSGSELEDPVV 303

Query: 294 NQIYKEFEVTANILYGNIEILDGIPVGEMVVILSGDEE 331
           + I K+F+VT NI YGN+E+L G P G +V+ L G  E
Sbjct: 304 SDITKKFDVTTNIFYGNVELLQGQPFGSLVLTLKGSSE 341
```

There is also homology to SEQ ID 76.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1826

A DNA sequence (GBSx1933) was identified in *S.agalactiae* <SEQ ID 5669> which encodes the amino acid sequence <SEQ ID 5670>. This protein is predicted to be ABC transporter, permease protein. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-12.79    Transmembrane  203 - 219 ( 197 - 225)
   INTEGRAL    Likelihood = -8.86    Transmembrane   73 -  89 (  69 - 102)
   INTEGRAL    Likelihood = -7.38    Transmembrane   38 -  54 (  35 -  56)
   INTEGRAL    Likelihood = -1.12    Transmembrane  103 - 119 ( 103 - 119)

----- Final Results -----
               bacterial membrane --- Certainty=0.6116(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10083> which encodes amino acid sequence <SEQ ID 10084> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB59829 GB:AJ012388 hypothetical protein [Lactococcus lactis]
 Identities = 137/231 (59%), Positives = 171/231 (73%), Gaps = 1/231 (0%)

Query: 1   MIEWIQTHLPNVYQMGWEGAYGWQTAIVQTLYMTFWSFLIGGLMGLLGGLFLVLTSPRGV 60
           M EW     PNV  +GW G  GW TAIVQTLYMTF S LIGGL+GL+ G+ +V+T+  G+
Sbjct: 1   MAEWFAHTFPNVVYLGWTGETGWWTAIVQTLYMTFISALIGGLLGLIFGIGVVVTAEDGI 60

Query: 61  IANKLVFGVLDKVVSVFRALPFIILLALIAPVTRVIVGTTLGSPAALVPLSLAVFPFFAR 120
             N+ +F +LDK+VS+ RA PFIILLA IAP+T+++VGT +G  AALVPL+L V PF+AR
Sbjct: 61  TPNRPLFWILDKIVSIGRAFPFIILLAAIAPLTKILVGTQIGVTAALVPLALGVAPFYAR 120

Query: 121 QVQVVLAELDGGVIEAAQASGGTLWDII-VVYLREGLPDLIRVSTVTLISLVGETAMAGA 179
           QVQ  L  +D G +EAAQ  G    DI+  VYLRE L  LIRVSTVTLISL+G TAMAGA
Sbjct: 121 QVQASLESVDHGKVEAAQTVGADFLDIVFTVYLREELASLIRVSTVTLISLIGLTAMAGA 180

Query: 180 IGAGGLGSVAITKGYNYSRDDITLVATILILLLIFFIQFLGDFLTRRLSHK 230
           IGAGGLG+ AI+ GYN   +D+T  ATILIL+ +  +Q +GDFL RR+SH+
Sbjct: 181 IGAGGLGNTAISYGYNRFANDVTWFATILILIFVLLVQLVGDFLARRVSHR 231
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5671> which encodes the amino acid sequence <SEQ ID 5672>. Analysis of this protein sequence reveals the following:

```
        Possible site: 32
   >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood =-11.15    Transmembrane  194 - 210 ( 187 - 215)
      INTEGRAL    Likelihood =-10.67    Transmembrane   28 -  44 (  20 -  52)
      INTEGRAL    Likelihood = -8.12    Transmembrane   70 -  86 (  62 -  91)

   ----- Final Results -----
                 bacterial membrane --- Certainty=0.5458(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB59829 GB:AJ012388 hypothetical protein [Lactococcus lactis]
 Identities = 123/213 (57%), Positives = 153/213 (71%), Gaps = 1/213 (0%)

Query: 9   GDAGWGLAIWNTLYMTIVPFIVGGAIGLLLGLLLVLTGPDGVIENKTICWVIDKVTSIFR 68
           G+ GW  AI  TLYMT +  ++GG +GL+ G+ +V+T  DG+  N+ + W++DK+ SI R
Sbjct: 19  GETGWWTAIVQTLYMTFISALIGGLLGLIFGIGVVVTAEDGITPNRPLFWILDKIVSIGR 78

Query: 69  AIPFVILIAILASFTYLLLRTTLGATAALVPLTFATFPFYARQVQVVFSELDKGVIEAAQ 128
           A PF+IL+A +A  T +L+ T +G TAALVPL     PFYARQVQ     +D G +EAAQ
Sbjct: 79  AFPFIILLAAIAPLTKILVGTQIGVTAALVPLALGVAPFYARQVQASLESVDHGKVEAAQ 138

Query: 129 ASGATFWDIV-KVYLSEGLPDLIRVSTVTLISLVGETAMAGAIGAGGLGNVAISYGYNRF 187
             GA F DIV  VYL E L  LIRVSTVTLISL+G TAMAGAIGAGGLGN AISYGYNRF
Sbjct: 139 TVGADFLDIVFTVYLREELASLIRVSTVTLISLIGLTAMAGAIGAGGLGNTAISYGYNRF 198

Query: 188 NNDVTWVATIIILLIIFAIQFIGDSLTRRFSHK 220
            NDVTW ATI+IL+ +  +Q +GD L RR SH+
Sbjct: 199 ANDVTWFATILILIFVLLVQLVGDFLARRVSHR 231
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 146/212 (68%), Positives = 172/212 (80%)

Query: 19  GAYGWQTAIVQTLYMTFWSFLIGGLMGLLGGLFLVLTSPRGVIANKLVFGVLDKVVSVFR 78
           G  GW  AI  TLYMT   F++GG +GLL GL LVLT P GVI NK +  V+DKV S+FR
Sbjct: 9   GDAGWGLAIWNTLYMTIVPFIVGGAIGLLLGLLLVLTGPDGVIENKTICWVIDKVTSIFR 68

Query: 79  ALPFIILLALIAPVTRVIVGTTLGSPAALVPLSLAVFPFFARQVQVVLAELDGGVIEAAQ 138
           A+PF+IL+A++A  T +++ TTLG+ AALVPL+ A FPF+ARQVQVV +ELD GVIEAAQ
Sbjct: 69  AIPFVILIAILASFTYLLLRTTLGATAALVPLTFATFPFYARQVQVVFSELDKGVIEAAQ 128

Query: 139 ASGGTLWDIIVVYLREGLPDLIRVSTVTLISLVGETAMAGAIGAGGLGSVAITKGYNYSR 198
           ASG T WDI+ VYL EGLPDLIRVSTVTLISLVGETAMAGAIGAGGLG+VAI+ GYN
Sbjct: 129 ASGATFWDIVKVYLSEGLPDLIRVSTVTLISLVGETAMAGAIGAGGLGNVAISYGYNRFN 188

Query: 199 DDITLVATILILLLIFFIQFLGDFLTRRLSHK 230
           +D+T VATI+ILL+IF IQF+GD LTRR SHK
Sbjct: 189 NDVTWVATIIILLIIFAIQFIGDSLTRRFSHK 220
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1827

A DNA sequence (GBSx1934) was identified in *S.agalactiae* <SEQ ID 5673> which encodes the amino acid sequence <SEQ ID 5674>. This protein is predicted to be alcohol dehydrogenase, zinc-containing (Zn-dependent). Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.92    Transmembrane   71 -  87 (  69 -  87)

----- Final Results -----
               bacterial membrane --- Certainty=0.2168(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9419> which encodes amino acid sequence <SEQ ID 9420> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF41759 GB:AE002488 alcohol dehydrogenase, zinc-containing
          [Neisseria meningitidis MC58]
 Identities = 135/246 (54%), Positives = 186/246 (74%), Gaps = 1/246 (0%)

Query: 3   SHCEDGGWILGHLIEGTQAEYVHIPHADGSLYHAPEGVCDDALVMLSDILPTSYEIGVLP 62
           SHC +GGWILG++I+GTQAEYV  P+AD SL   P+ V ++  ++LSD LPT++EIGV
Sbjct: 102 SHCRNGGWILGYMIDGTQAEYVRTPYADNSLVPLPDNVNEEIALLLSDALPTAHEIGVQY 161

Query: 63  SHIKPGDTVCIVGAGPIGLSALLTAQFYSPAKIIMVDLSQKRLEASKKFGATHTILSTST 122
             +KPGDTV I GAGP+G+SALLTAQ YSPA II+ D+ + RL+ +K+ GATHTI + ++
Sbjct: 162 GDVKPGDTVFIAGAGPVGMSALLTAQLYSPAAIIVCDMDENRLKLAKELGATHTI-NPAS 220

Query: 123 QEVKEEIDKITKGRGVDVVLECVGYPATFDICQNVVSIGGHIANVGVHGKPVEFNLQDLW 182
            EV +++  I    GVD  +E VG PAT+++CQ++V  GGHIA VGVHG+ V+F L+ LW
Sbjct: 221 GEVSKQVFAIVGEDGVDCAIEAVGIPATWNMCQDIVKPGGHIAVVGVHGQSVDFKLEKLW 280

Query: 183 IKNITLNTGLVNANTTEMLLEVLETGKIDATQLVTHHFKLSEIEEAYKVFKAAEENNTLK 242
           IK + + TGLVNANTTEML++ + +  +D T+++THHFK SE+E+AY VFK A EN  +K
Sbjct: 281 IKKLAITTGLVNANTTEMLMKAISSSSVDYTKMLTHHFKFSELEKAYDVFKHAAENQVMK 340

Query: 243 VIIEND 248
           V++E D
Sbjct: 341 VVLEAD 346
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 785> which encodes the amino acid sequence <SEQ ID 786>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -5.41    Transmembrane  184 - 200 ( 183 - 203)

----- Final Results -----
               bacterial membrane --- Certainty=0.3166(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 199/250 (79%), Positives = 226/250 (89%)

Query: 1   MPSHCEDGGWILGHLIEGTQAEYVHIPHADGSLYHAPEGVCDDALVMLSDILPTSYEIGV 60
           + SHC+DGGWILGHLI GTQAEYVHIPHADGSLYHAP+ + D+ALVMLSDILPTSYEIGV
Sbjct: 114 LSSHCQDGGWILGHLINGTQAEYVHIPHADGSLYHAPDTIDDEALVMLSDILPTSYEIGV 173

Query: 61  LPSHIKPGDTVCIVGAGPIGLSALLTAQFYSPAKIIMVDLSQKRLEASKKFGATHTILST 120
           LPSH+KPGD VCIVGAGP+GL+ALLT QF+SPA IIMVDLSQ RLEA+K FGATHTI S
Sbjct: 174 LPSHVKPGDNVCIVGAGPVGLAALLTVQFFSPANIIMVDLSQNRLEAAKTFGATHTICSG 233

Query: 121 STQEVKEEIDKITKGRGVDVVLECVGYPATFDICQNVVSIGGHIANVGVHGKPVEFNLQD 180
           S++EVK  ID IT GRGVD+ +ECVGYPATFDICQ ++S+GGHIANVGVHGKPV+FNL +
Sbjct: 234 SSEEVKAIIDDITNGRGVDISMECVGYPATFDICQKIISVGGHIANVGVHGKPVDFNLDE 293
```

```
Query: 181 LWIKNITLNTGLVNANTTEMLLEVLETGKIDATQLVTHHFKLSEIEEAYKVFKAAEENNT 240
           LWIKNITLNTGLVNANTTEMLL VL+TGKIDAT+L+THHFKLSE+E+AY+ FK A  NN
Sbjct: 294 LWIKNITLNTGLVNANTTEMLLNVLKTGKIDATRLITHHFKLSEVEKAYETFKHAGANNA 353

Query: 241 LKVIIENDIT 250
           LKVII+NDI+
Sbjct: 354 LKVIIDNDIS 363
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1828

A DNA sequence (GBSx1935) was identified in *S.agalactiae* <SEQ ID 5675> which encodes the amino acid sequence <SEQ ID 5676>. This protein is predicted to be a dehydrogenase fragment. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.46   Transmembrane   47 -  63 (  33 -  66)

----- Final Results -----
               bacterial membrane --- Certainty=0.5182(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

There is also homology to SEQ ID 786:

```
Identities = 23/38 (60%), Positives = 28/38 (73%)

Query: 7  WRNSNMRAATYLSANELSLTDKAKPQVIKPTDAVVXLV 44
          ++  NM+AATYLS    L L DK KP +IKPTDA+V LV
Sbjct: 10 YKKLNMKAATYLSTGNLQLIDKPKPVIIKPTDAIVQLV 47
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1829

A DNA sequence (GBSx1936) was identified in *S.agalactiae* <SEQ ID 5677> which encodes the amino acid sequence <SEQ ID 5678>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1001(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1830

A DNA sequence (GBSx1937) was identified in *S.agalactiae* <SEQ ID 5679> which encodes the amino acid sequence <SEQ ID 5680>. This protein is predicted to be branched chain amino acid transport system II carrier protein (brnQ). Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.66    Transmembrane  158 - 174 ( 154 - 177)
   INTEGRAL    Likelihood = -6.64    Transmembrane  233 - 249 ( 231 - 252)
   INTEGRAL    Likelihood = -5.20    Transmembrane   37 -  53 (  30 -  57)
   INTEGRAL    Likelihood = -3.98    Transmembrane   90 - 106 (  87 - 108)
   INTEGRAL    Likelihood = -0.80    Transmembrane  130 - 146 ( 130 - 146)

----- Final Results -----
               bacterial membrane --- Certainty=0.4864(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9417> which encodes amino acid sequence <SEQ ID 9418> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC00400 GB:AF008220 branch-chain amino acid transporter
          [Bacillus subtilis]
 Identities = 89/250 (35%), Positives = 139/250 (55%), Gaps = 18/250 (7%)

Query: 1   MDALASIAFAIIVIQASKQYGAITKKEITSMALKSGAIATFLLAFIYIFVGRIGATSQSL 60
           MDALASI F ++V+ A K  G    K + +  +K+G IA   L FIY+ +  +GATS  +
Sbjct: 199 MDALASIVFGVVVVNAVKSKGVTQSKALAAACIKAGVIAALGLTFIYVSLAYLGATSTNA 258

Query: 61  FKFANGSFLLHNTPI-DGGHVLSQSANFYLGIVGQAILGTAIFLACLTTATGLITACAEY 119
                        P+ +G  +LS S+++  G +G  +LG AI +ACLTT+ GL+T+C +Y
Sbjct: 259 IG-----------PVGEGAKILSASSHYLFGSLGNIVLGAAITVACLTTSIGLVTSCGQY 307

Query: 120 FHKLLPKISHITWATIFTLIAITFYFGGLSEIIRWSLPVLYLLYPLTIVLIFLVFFDQKF 179
           F KL+P +S+    TI TL ++     GL++II +S+P+L  +YPL IV+I L F D+ F
Sbjct: 308 FSKLIPALSYKIVVTIVTLFSLIIANFGLAQIIAFSVPILSAIYPLAIVIIVLSFIDKIF 367

Query: 180 ESSRIVYQTSIAATAVAALYDALSKLGEMTGLFTIPSALTTFFTKVVPLGEYSMGWISFA 239
           +  R VY   +  T + ++ D +   G   G      +L  F    +PL    +GW+
Sbjct: 368 KERREVYIACLIGTGLFSILDGIKAAGFSLG------SLDVFLNANLPLYSLGIGWVLPG 421

Query: 240 ICGVLVGLIL 249
           I G ++G +L
Sbjct: 422 IVGAVIGYVL 431
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2233> which encodes the amino acid sequence <SEQ ID 2234>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-10.83    Transmembrane  235 - 251 ( 228 - 258)
   INTEGRAL    Likelihood = -8.49    Transmembrane  434 - 450 ( 429 - 454)
   INTEGRAL    Likelihood = -8.12    Transmembrane  359 - 375 ( 356 - 377)
   INTEGRAL    Likelihood = -7.86    Transmembrane  150 - 166 ( 144 - 171)
   INTEGRAL    Likelihood = -6.00    Transmembrane  298 - 314 ( 288 - 316)
   INTEGRAL    Likelihood = -5.95    Transmembrane   42 -  58 (  38 -  63)
   INTEGRAL    Likelihood = -3.35    Transmembrane  336 - 352 ( 335 - 354)
   INTEGRAL    Likelihood = -2.81    Transmembrane  199 - 215 ( 198 - 218)
   INTEGRAL    Likelihood = -2.18    Transmembrane  120 - 136 ( 120 - 138)
   INTEGRAL    Likelihood = -1.81    Transmembrane  390 - 406 ( 390 - 407)
   INTEGRAL    Likelihood = -1.01    Transmembrane   81 -  97 (  81 -  97)
```

```
----- Final Results -----
                bacterial membrane --- Certainty=0.5331(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 161/253 (63%), Positives = 197/253 (77%)

Query: 1   MDALASIAFAIIVIQASKQYGAITKKEITSMALKSGAIATFLLAFIYIFVGRIGATSQSL 60
           MDALAS+ FAI+VI+A+KQ+GA T KE+T + L SGAIA  LLA +YIFVGRIGATSQSL
Sbjct: 202 MDALASLVFAILVIEATKQFGAKTDKEMTKITLISGAIAILLLALVYIFVGRIGATSQSL 261

Query: 61  FKFANGSFLLHNTPIDGGHVLSQSANFYLGIVGQAILGTAIFLACLTTATGLITACAEYF 120
           F F +GSF LH  P++GG +LS ++ FYLG +GQA L    IFLACLTT+TGLIT+ AEYF
Sbjct: 262 FPFIDGSFTLHGNPVNGGQILSHASRFYLGGIGQAFLAVVIFLACLTTSTGLITSSAEYF 321

Query: 121 HKLLPKISHITWATIFTLIAITFYFGGLSEIIRWSLPVLYLLYPLTIVLIFLVFFDQKFE 180
           HKL+P +SHI WATIFTL++  FYFGGLS II WS PVL+LLYPLT+ LIFLV   + F
Sbjct: 322 HKLVPALSHIAWATIFTLLSAFFYFGGLSVIINWSAPVLFLLYPLTVDLIFLVLAQKCFN 381

Query: 181 SSRIVYQTSIAATAVAALYDALSKLGEMTGLFTIPSALTTFFTKVVPLGEYSMGWISFAI 240
           +  IVY+T+I  T + A++DAL  L +MTGLF +P A+ TFF K VPLG++SMGWI FA
Sbjct: 382 NDPIVYRTTIGLTFIPAIFDALLTLSQMTGLFHLPEAVVTFFQKTVPLGQFSMGWIIFAA 441

Query: 241 CGVLVGLILKKVK 253
            G L+GLIL K K
Sbjct: 442 IGFLIGLILSKTK 454
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1831

A DNA sequence (GBSx1938) was identified in *S.agalactiae* <SEQ ID 5681> which encodes the amino acid sequence <SEQ ID 5682>. This protein is predicted to be 30S ribosomal protein S12 (rpsL). Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3698(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9429> which encodes amino acid sequence <SEQ ID 9430> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA78825 GB:Z15120 ribosomal protein S12 [Streptococcus pneumoniae]
 Identities = 64/71 (90%), Positives = 68/71 (95%)

Query: 1  MPTINQLVRKPRKSKVEKSDSPALNIGYNSHRKVHTKLSAPQKRGVATRVGTMTPKKPNS 60
          MPTINQLVRKPRKSKVEKS SPALN+GYNSH+KV T +S+PQKRGVATRVGTMTPKKPNS
Sbjct: 1  MPTINQLVRKPRKSKVEKSKSPALNVGYNSHKKVQTNVSSPQKRGVATRVGTMTPKKPNS 60

Query: 61 ALRKFARVRLS 71
          ALRKFARVRLS
Sbjct: 61 ALRKFARVRLS 71
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5683> which encodes the amino acid sequence <SEQ ID 5684>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3879(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 44/48 (91%), Positives = 47/48 (97%)

Query: 24 LNIGYNSHRKVHTKLSAPQKRGVATRVGTMTPKKPNSALRKFARVRLS 71
          LNIGYNSH+KV TK++APQKRGVATRVGTMTPKKPNSALRKFARVRLS
Sbjct: 1  LNIGYNSHKKVQTKMAAPQKRGVATRVGTMTPKKPNSALRKFARVRLS 48
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1832

A DNA sequence (GBSx1939) was identified in *S.agalactiae* <SEQ ID 5685> which encodes the amino acid sequence <SEQ ID 5686>. This protein is predicted to be purR. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -0.37   Transmembrane  142 - 158 ( 142 - 159)

----- Final Results -----
               bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA10902 GB:AJ222642 purR [Lactococcus lactis]
 Identities = 143/269 (53%), Positives = 195/269 (72%), Gaps = 1/269 (0%)

Query: 3   LRRSERMVVISNYLINNPYTLTSLNTFASKYGAAKSSISEDIAIIKKAFEQAQIGDIKTV 62
           ++R+ER+V  +N+LIN+P  + +LN  +  Y  AKSSISED+  IK+ FE   +G ++T
Sbjct: 1   MKRNERLVDFTNFLINHPNQMLNLNELSKHYEVAKSSISEDLVFIKRVFENQGVGLVETF 60

Query: 63  TGASGGVIFTPTIAEAEAKEIVEELRQRLSENDRILPGGYIYLSDLLSTPKMLQSIGRII 122
            G+ GGV FTP I +  + E+ +E+ + L E +RILPGGYIYLSD+L TP  L+ IG+II
Sbjct: 61  PGSLGGVRFTPYITDERSLEMSQEIAELLREENRILPGGYIYLSDILGTPSNLRKIGQII 120

Query: 123 ANAYRGQKIDAVMTVATKGVPLANAVANVLDVPFVIVRRDLKITEGSTVSVNYASGSSGR 182
           A+ Y  +++D VMT+ATKG+P+A +VA +LDVPFVIVRRD K+TEG+T++VNY SGSS R
Sbjct: 121 AHEYHEKQVDVVMTIATKGIPIAQSVAEILDVPFVIVRRDPKVTEGATLNVNYMSGSSSR 180

Query: 183 IEKMFLSKRSLKPNSRVLIVDDFLKGGGTVSGMISLLSEFDSTLVGVAVFAENA-QEQRE 241
           +E M LSKRSL     VLIVDDF+KG GT++GM SL+ EFD  L GVAVF E   + +R
Sbjct: 181 VENMTLSKRSLSIGQNVLIVDDFMKGAGTINGMRSLVHEFDCLLAGVAVFLEGPFKGERL 240

Query: 242 KMAYKSLLRVSEIDVKNNRVSVEAGNIFD 270
              YKS+L+V  ID+ N  + V+ GNIF+
Sbjct: 241 IDDYKSILKVDRIDIANRSIDVQLGNIFN 269
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5687> which encodes the amino acid sequence <SEQ ID 5688>. Analysis of this protein sequence reveals the following:

```
        Possible site: 41
   >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -1.97   Transmembrane  142 - 158 ( 142 - 160)

   ----- Final Results -----
                bacterial membrane --- Certainty=0.1786(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA10902 GB:AJ222642 purR [Lactococcus lactis]
 Identities = 142/269 (52%), Positives = 196/269 (72%), Gaps = 1/269 (0%)

Query: 3   LRRSERMVVISNYLINNPYKLTSLNTFATKYEAAKSSISEDIAIIKKAFEEANIGDIDTL 62
           ++R+ER+V  +N+LIN+P ++ +LN  +  YE AKSSISED+  IK+ FE   +G ++T
Sbjct: 1   MKRNERLVDFTNFLINHPNQMLNLNELSKHYEVAKSSISEDLVFIKRVFENQGVGLVETF 60

Query: 63  TGASGGVIFTPSISETEARTIVEDLCQRLSESDRILPGGYIYLSDLLSTPKILQNIGRII 122
            G+ GGV FTP I++  +  + +++ + L E +RILPGGYIYLSD+L TP  L+ IG+II
Sbjct: 61  PGSLGGVRFTPYITDERSLEMSQEIAELLREENRILPGGYIYLSDILGTPSNLRKIGQII 120

Query: 123 ANAFKGEKIDAVMTVATKGVPLANAVANILSVPFVIVRRDLKITEGSTVSVNYASASSDR 182
           A+ +  +++D VMT+ATKG+P+A +VA IL VPFVIVRRD K+TEG+T++VNY S SS R
Sbjct: 121 AHEYHEKQVDVVMTIATKGIPIAQSVAEILDVPFVIVRRDPKVTEGATLNVNYMSGSSSR 180

Query: 183 IEKMFLSKRSLKPNSRVLIVDDFLKGGGTITGMISLLTEFDSTLVGVAVFAENA-QSERE 241
           +E M LSKRSL     VLIVDDF+KG GTI GM SL+ EFD  L GVAVF E   + ER
Sbjct: 181 VENMTLSKRSLSIGQNVLIVDDFMKGAGTINGMRSLVHEFDCLLAGVAVFLEGPFKGERL 240

Query: 242 QMTFKSLLKVSEIDVKNNNVVVEVGNIFD 270
              +KS+LKV  ID+ N ++ V++GNIF+
Sbjct: 241 IDDYKSILKVDRIDIANRSIDVQLGNIFN 269
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 234/270 (86%), Positives = 255/270 (93%)

Query: 1   MKLRRSERMVVISNYLINNPYTLTSLNTFASKYGAAKSSISEDIAIIKKAFEQAQIGDIK 60
           MKLRRSERMVVISNYLINNPY LTSLNTFA+KY AAKSSISEDIAIIKKAFE+A IGDI
Sbjct: 1   MKLRRSERMVVISNYLINNPYKLTSLNTFATKYEAAKSSISEDIAIIKKAFEEANIGDID 60

Query: 61  TVTGASGGVIFTPTIAEAEAKEIVEELRQRLSENDRILPGGYIYLSDLLSTPKMLQSIGR 120
           T+TGASGGVIFTP+I+E EA+ IVE+L QRLSE+DRILPGGYIYLSDLLSTPK+LQ+IGR
Sbjct: 61  TLTGASGGVIFTPSISETEARTIVEDLCQRLSESDRILPGGYIYLSDLLSTPKILQNIGR 120

Query: 121 IIANAYRGQKIDAVMTVATKGVPLANAVANVLDVPFVIVRRDLKITEGSTVSVNYASGSS 180
           IIANA++G+KIDAVMTVATKGVPLANAVAN+L VPFVIVRRDLKITEGSTVSVNYAS SS
Sbjct: 121 IIANAFKGEKIDAVMTVATKGVPLANAVANILSVPFVIVRRDLKITEGSTVSVNYASASS 180

Query: 181 GRIEKMFLSKRSLKPNSRVLIVDDFLKGGGTVSGMISLLSEFDSTLVGVAVFAENAQEQR 240
            RIEKMFLSKRSLKPNSRVLIVDDFLKGGGT++GMISLL+EFDSTLVGVAVFAENAQ +R
Sbjct: 181 DRIEKMFLSKRSLKPNSRVLIVDDFLKGGGTITGMISLLTEFDSTLVGVAVFAENAQSER 240

Query: 241 EKMAYKSLLRVSEIDVKNNRVSVEAGNIFD 270
           E+M +KSLL+VSEIDVKNN V VE GNIFD
Sbjct: 241 EQMTFKSLLKVSEIDVKNNNVVVEVGNIFD 270
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1833

A DNA sequence (GBSx1940) was identified in *S.agalactiae* <SEQ ID 5689> which encodes the amino acid sequence <SEQ ID 5690>. This protein is predicted to be cmp-binding-factor 1. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1753(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC44803 GB:U21636 cmp-binding-factor 1 [Staphylococcus aureus]
 Identities = 140/310 (45%), Positives = 195/310 (62%), Gaps = 6/310 (1%)

Query: 3   INQMKKDELFEGFYLIKKAEVRKTRAGKDFIAFTFQDDTGEISGNMWDAQTYNVEEFVAG 62
           I  +   +  + F+L+ KA    T  GKD++    QD +GEI    W A   ++
Sbjct: 4   IENLNPGDSVDHFFLVHKATQGVTAQGKDYMTLHLQDKSGEIEAKFWTATKNDMATIKPE 63

Query: 63  KIVHMKGRREVYNGTPQ--VNQITLRNIKDGEPNDPRDFKEKPPINVDNVREYMEQMLFK 120
           +IVH+KG    Y G  Q  VNQI L   +D    +   F +  P++   ++E +   L
Sbjct: 64  EIVHVKGDIINYRGNKQMKVNQIRLATTEDQLKTE--QFVDGAPLSPAEIQEEISHYLLD 121

Query: 121 IENATWQRVVRALYRKYNKEFFTYPAAKTNHHAFESGLAYHTATMVRLADSIGDIYPELN 180
           IENA  QR+ R L +KY + F+TYPAA ++HH F SGL+YH  TM+R+A SI DIYP LN
Sbjct: 122 IENANLQRITRHLLKKYQERFYTYPAASSHHHNFASGLSYHVLTMLRIAKSICDIYPLLN 181

Query: 181 KSLMFAGIMLHDLAKVIELSGPDNTEYTIRGNLIGHISLIDEELTKILAELNIDDTKEEV 240
           KSL+++GI+LHD+ KV ELSGP  T YT+ GNL+GHIS+  +E+ +   ELNI+   EE+
Sbjct: 182 KSLLYSGIILHDIGKVRELSGPVATSYTVEGNLLGHISIASDEVVEAARELNIEG--EEI 239

Query: 241 TVLRHVILSHHGQLEYGSPVRPRIMEAEIIHMIDNIDANMMMMTTALNRVNEGEMTNRIF 300
            +LRH+ILSHHG+LEYGSP  P + EAEI+  IDNIDA M M   A  + ++G+ T++IF
Sbjct: 240 MLLRHMILSHHGKLEYGSPKLPYLKEAEILCYIDNIDARMNMFEKAYKKTDKGQFTDKIF 299

Query: 301 AMDNRSFYKP 310
            ++NR FY P
Sbjct: 300 GLENRRFYNP 309
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5691> which encodes the amino acid sequence <SEQ ID 5692>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1822(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 275/311 (88%), Positives = 300/311 (96%)

Query: 1   MKINQMKKDELFEGFYLIKKAEVRKTRAGKDFIAFTFQDDTGEISGNMWDAQTYNVEEFV 60
           MKINQMKKD+LFEGFYLIK AEVRKTRAGKDFI+ TFQDDTGEISGN+WDAQ YNVEEF
Sbjct: 1   MKINQMKKDQLFEGFYLIKSAEVRKTRAGKDFISLTFQDDTGEISGNLWDAQPYNVEEFT 60

Query: 61  AGKIVHMKGRREVYNGTPQVNQITLRNIKDGEPNDPRDFKEKPPINVDNVREYMEQMLFK 120
           AGK+V MKGRREVYNGTPQVNQITLRN++ GEPNDP+DFKEK P++V  VR+Y+EQMLFK
Sbjct: 61  AGKVVFMKGRREVYNGTPQVNQITLRNVRPGEPNDPKDFKEKAPVSVTEVRDYLEQMLFK 120
```

```
Query: 121 IENATWQRVVRALYRKYNKEFFTYPAAKTNHHAFESGLAYHTATMVRLADSIGDIYPELN 180
           IENATWQR+VRALYRKY+KEF+TYPAAKTNHHAFESGLAYHTATMVRLADSIGDIYP+LN
Sbjct: 121 IENATWQRIVRALYRKYDKEFYTYPAAKTNHHAFESGLAYHTATMVRLADSIGDIYPDLN 180

Query: 181 KSLMFAGIMLHDLAKVIELSGPDNTEYTIRGNLIGHISLIDEELTKILAELNIDDTKEEV 240
           KSL+FAGIMLHDLAKVIEL+GPDNTEYT+RGNLIGHISLI+EE+TK+++EL IDDTKEEV
Sbjct: 181 KSLLFAGIMLHDLAKVIELTGPDNTEYTVRGNLIGHISLINEEITKVISELQIDDTKEEV 240

Query: 241 TVLRHVILSHHGQLEYGSPVRPRIMEAEIIHMIDNIDANMMMMTTALNRVNEGEMTNRIF 300
            VLRHVILSHHGQLEYGSPVRPRIMEAEIIHMIDNIDANMMMMTTAL+RV+EGEMTNRIF
Sbjct: 241 IVLRHVILSHHGQLEYGSPVRPRIMEAEIIHMIDNIDANMMMMTTALSRVSEGEMTNRIF 300

Query: 301 AMDNRSFYKPN 311
           AMDNRSFYKPN
Sbjct: 301 AMDNRSFYKPN 311
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1834

A DNA sequence (GBSx1941) was identified in *S.agalactiae* <SEQ ID 5693> which encodes the amino acid sequence <SEQ ID 5694>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-14.59    Transmembrane    2 -  18 (   1 -  22)

----- Final Results -----
               bacterial membrane --- Certainty=0.6838(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5695> which encodes the amino acid sequence <SEQ ID 5696>. Analysis of this protein sequence reveals the following:

```
     Possible site: 17

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.05    Transmembrane    3 -  19 (   1 -  26)

----- Final Results -----
               bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 309/424 (72%), Positives = 370/424 (86%), Gaps = 3/424 (0%)

Query: 1   MLVIILIIVLASLTVTIISYQKMTELTKSVEKQLEDNADNLSDQLTYQIEVAQKDQILTL 60
           +++ +L++VL  L   ++   K+  L   + + LE NADNLSDQ+TYQ++ A K Q+L L
Sbjct: 3   LILFLLVLVLLGLGAYLLF--KVNGLQHQLAQTLEGNADNLSDQMTYQLDTANKQQLLEL 60

Query: 61  TNQLNRMQQEIYQLLTDMRTELNQHLTESRDRSDKRLELINSNLSQSVQKMQDSNEKRLD 120
           T  +NR Q  +YQ LTD+R  L++ L++SRDRSDKRLE IN  ++QS++ MQ+SNEKRL+
Sbjct: 61  TQLMNRQQAGLYQQLTDIRDVLHRSLSDSRDRSDKRLEKINQQVNQSLKNMQESNEKRLE 120

Query: 121 QMRQTVEEKLEKTLQTRLQTSFETVSRQLESVNQGLGEMKTVAQDVGTLNKVLSNTKTRG 180
           +MRQ VEEKLE+TL+ RL  SF++VS+QLESVN+GLGEM++VAQDVGTLNKVLSNTKTRG
Sbjct: 121 KMRQIVEEKLEETLKNRLHASFDSVSKQLESVNKGLGEMRSVAQDVGTLNKVLSNTKTRG 180

Query: 181 ILGELQLGQIIEDIMTVSQYEREFPTVSGSSERVEYAIKLPGNGQGDYIYLPIDSKFPLE 240
           ILGELQLGQIIEDIMT SQYEREF TVSGSSERVEYAIKLPGNGQG YIYLPIDSKFPLE
Sbjct: 181 ILGELQLGQIIEDIMTSSQYEREFVTVSGSSERVEYAIKLPGNGQGGYIYLPIDSKFPLE 240
```

```
Query: 241 DYYRLEDAYELGDKVQIELYRKSLLASIRKFAKDINNKYLNPPETTNFGIMFLPTEGLYS 300
           DYYRLEDAYE+GDK+ IE  RK+LLA+I++FAKDI+ KYLNPPETTNFG+MFLPTEGLYS
Sbjct: 241 DYYRLEDAYEVGDKLAIEASRKALLAAIKRFAKDIHKKYLNPPETTNFGVMFLPTEGLYS 300

Query: 301 EVVRNATFFDSLRRDENIVVAGPSTLSALLNSLSVGFKTLNIQKNANDISKILGNVKVEF 360
           EVVRNA+FFDSLRR+ENIVVAGPSTLSALLNSLSVGFKTLNIQKNA+DISKILGNVK+EF
Sbjct: 301 EVVRNASFFDSLRREENIVVAGPSTLSALLNSLSVGFKTLNIQKNADDISKILGNVKLEF 360

Query: 361 GKFGGMLSKAQKQLNTASKSIDSLLTTRTNAIIRVLNTVEEHQDQATTSLLNLPITEEEE 420
            KFGG+L+KAQKQ+NTA+ ++D L++TRTNAI+R LNTVE +QDQAT SLLN+P+ EEE
Sbjct: 361 DKFGGLLAKAQKQMNTANNTLDQLISTRTNAIVRALNTVETYQDQATKSLLNMPLLEEEN 420

Query: 421 INEN 424
            NEN
Sbjct: 421 -NEN 423
```

SEQ ID 5694 (GBS88) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 18 (lane 2; MW 48kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1835

A DNA sequence (GBSx1942) was identified in *S.agalactiae* <SEQ ID 5697> which encodes the amino acid sequence <SEQ ID 5698>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2722(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB13453 GB:Z99112 yloS [Bacillus subtilis]
 Identities = 75/217 (34%), Positives = 109/217 (49%), Gaps = 12/217 (5%)

Query: 1    MTKIALFAGG------DLTYFEYDFDYFVGIDRGSLFLLKNGLSLDMAVGDFDSITEDEL 54
            M  I + AGG      DLT +  +   ++G+D+G++ LL  G+    A GDFDSITE E
Sbjct: 1    MKTINIVAGGPKNLIPDLTGYTDEHTLWIGVDKGTVTLLDAGIIPVEAFGDFDSITEQER 60

Query: 55   LYIKHYCSNIVSASAEKNDTDTELALKTIFKEFPEAQVTVFGAFGGRIDHMMSNIFLPSD 114
              I+     +    AEK+ TD +LAL    ++ P+  + +FG  GGR DH + NI L
Sbjct: 61   RRIEKAAPALHVYQAEKDQTDLDLALDWALEKQPDI-IQIFGITGGRADHFLGNIQLLYK 119

Query: 115 RDLEPFMSQIRLKDEQNIVTYLPSGKNQVSRIEGMSYVSFMPESES--TLQISGAKYELN 172
                   +IRL D+QN +   P G+  + + E   Y+SF+P SE    L ++G KY LN
Sbjct: 120 GVKTNI--KIRLIDKQNHIQMFPPGEYDIEKDENKRYISFIPFSEDIHELTLTGFKYPLN 177

Query: 173 KSNY-FKKKMYSSNEFMTSPIEVELKDGYLIIIYSKD 208
              +      + SNE + S        G LI+I S D
Sbjct: 178 NCHITLGSTLCISNELIHSRGTFSFAKGILIMIRSTD 214
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5699> which encodes the amino acid sequence <SEQ ID 5700>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
        bacterial cytoplasm --- Certainty=0.2467(Affirmative) < succ>
         bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 130/208 (62%), Positives = 166/208 (79%)

Query: 1   MTKIALFAGGDLTYFEYDFDYFVGIDRGSLFLLKNGLSLDMAVGDFDSITEDELLYIKHY 60
           M+K+ALFAGGDL+Y   DFDYFVGIDRGSLFLL+NGL L+MAVGDFDS+++     IK
Sbjct: 1   MSKVALFAGGDLSYISRDFDYFVGIDRGSLFLLENGLPLNMAVGDFDSVSQKAFTDIKEK 60

Query: 61  CSNIVSASAEKNDTDTELALKTIFKEFPEAQVTVFGAFGGRIDHMMSNIFLPSDRDLEPF 120
               ++A  EKNDTDTELALK +F  FPEA+VT+FGAFGGR+DH++SNIFLPSD  + PF
Sbjct: 61  AELFITAHPEKNDTDTELALKEVFARFPEAEVTIFGAFGGRMDHLLSNIFLPSDPGIAPF 120

Query: 121 MSQIRLKDEQNIVTYLPSGKNQVSRIEGMSYVSFMPESESTLQISGAKYELNKSNYFKKK 180
           M+QI L+D+QN++TY P+G++ + + EGM+YV+FM E E+ L I+GAK+EL + N+FKKK
Sbjct: 121 MAQIALRDQQNMITYRPAGQHLIHQEEGMTYVAFMAEGEADLTITGAKFELTQDNFFKKK 180

Query: 181 MYSSNEFMTSPIEVELKDGYLIIIYSKD 208
           +YSSN F+  PI V L  GYLIII SKD
Sbjct: 181 IYSSNAFIHQPITVSLPSGYLIIIQSKD 208
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1836

A DNA sequence (GBSx1943) was identified in *S.agalactiae* <SEQ ID 5701> which encodes the amino acid sequence <SEQ ID 5702>. This protein is predicted to be ribulose-phosphate 3-epimerase (rpe). Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.59   Transmembrane  124 - 140 ( 124 - 141)

----- Final Results -----
               bacterial membrane --- Certainty=0.1638(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06221 GB:AP001515 unknown conserved protein [Bacillus halodurans]
 Identities = 113/211 (53%), Positives = 153/211 (71%)

Query: 5   KIAPSILAADYANFANELKRIEETTAEYVHIDIMDGQFVPNISFGADVVSSMRKHSKLVF 64
           KIAPSIL+AD+AN  NE++ +E   A+Y+H+D+MDG FVPNI+ G  +V ++R  + L
Sbjct: 3   KIAPSILSADFANLGNEIQDVERGGADYIHVDVMDGHFVPNITIGPLIVDAIRPVTTLPL 62

Query: 65  DCHLMVVDPERYIEAFAQAGADIMTIHVEATKHIHGALQKIKEAGMKAGVVINPGTPVES 124
           D HLM+  P+ YI AFA+AGADI+T+HVEA  H+H  L  IKE+G+KAGVV+NP TPV S
Sbjct: 63  DVHLMIEQPDGYIPAFAKAGADIITVHVEACPHLHRTLHLIKESGVKAGVVLNPATPVSS 122

Query: 125 LIPILDLVDQILIMTVNPGFGGQAFIPEMMSKVKTVAAWRKEYGHHYDIEVDGGIDNTTI 184
           +  +L  VD +L MTVNPGFGGQ FIP ++ K+K +A+ +KE G  ++IEVDGG++  T
Sbjct: 123 IQHVLSDVDMVLFMTVNPGFGGQRFIPSVLPKLKELASLKKEQGLTFEIEVDGGVNEETA 182

Query: 185 KAAAEAGANVFVAGSYLFKASDLPAQVETLR 215
           K   EAGANV VAGS +F   D  A ++ +R
Sbjct: 183 KQCVEAGANVLVAGSAVFNEEDRAAAIKGIR 213
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5703> which encodes the amino acid sequence <SEQ ID 5704>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0072(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 183/219 (83%), Positives = 198/219 (89%)

Query: 1   MSTNKIAPSILAADYANFANELKRIEETTAEYVHIDIMDGQFVPNISFGADVVSSMRKHS 60
            MST KIAPSILAADYANFA+EL RIEET AEYVHIDIMDGQFVPNISFGADVV+SMRKHS
Sbjct: 1   MSTLKIAPSILAADYANFASELARIEETDAEYVHIDIMDGQFVPNISFGADVVASMRKHS 60

Query: 61  KLVFDCHLMVVDPERYIEAFAQAGADIMTIHVEATKHIHGALQKIKEAGMKAGVVINPGT 120
            KLVFDCHLMVVDPERY+EAFAQAGADIMTIH E+T+HIHGALQKIK AGMKAGVVINPGT
Sbjct: 61  KLVFDCHLMVVDPERYVEAFAQAGADIMTIHTESTRHIHGALQKIKAAGMKAGVVINPGT 120

Query: 121 PVESLIPILDLVDQILIMTVNPGFGGQAFIPEMMSKVKTVAAWRKEYGHHYDIEVDGGID 180
            P  +L P+LDLVDQ+LIMTVNPGFGGQAFIPE + KV TVA WR E G  +DIEVDGG+D
Sbjct: 121 PATALEPLLDLVDQVLIMTVNPGFGGQAFIPECLEKVATVAKWRDEKGLSFDIEVDGGVD 180

Query: 181 NTTIKAAAEAGANVFVAGSYLFKASDLPAQVETLRVALD 219
            N TI+A  EAGANVFVAGSYLFKASDL +QV+TLR AL+
Sbjct: 181 NKTIRACYEAGANVFVAGSYLFKASDLVSQVQTLRTALN 219
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1837

A DNA sequence (GBSx1944) was identified in *S.agalactiae* <SEQ ID 5705> which encodes the amino acid sequence <SEQ ID 5706>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2098(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB13451 GB:Z99112 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 148/296 (50%), Positives = 202/296 (68%), Gaps = 14/296 (4%)

Query: 2   QGRIVKSLAGFYYV----ESDGVVYQTRARGNFRKKGQIPYVGDWVEFSSQDQSEGYILS 57
            +G+I+K+L+GFYYV    E    V Q R RG FRK    P VGD+V + +++  EGY++
Sbjct: 3   EGKIIKALSGFYYVLDESEDSDKVIQCRGRGIFRKNKITPLVGDYVVYQAENDKEGYLME 62

Query: 58  IEERKNSLVRPPIVNIDQAVVIMSAKEPDFNANLLDRFLVLLEYKMIQPIIYISKLDLLD 117
            I+ER N L+RPPI N+DQAV++ SA +P F+  LLDRFLVL+E   IQPII I+K+DL++
Sbjct: 63  IKERTNELIRPPICNVDQAVLVFSAVQPSFSTALLDRFLVLVEANDIQPIICITKMDLIE 122

Query: 118 DLVVIDDIR---EHYQNIGY-VFCYSQEE------LLPLLANKVTVFMGQTGVGKSTLLN 167
            D    D I+   E Y+NIGY V+  S ++      ++P   +K TVF GQ+GVGKS+LLN
Sbjct: 123 DQDTEDTIQAYAEDYRNIGYDVYLTSSKDQDSLADIIPHFQDKTTVFAGQSGVGKSSLLN 182

Query: 168 KIAPELKLETGEISGSLGRGRHTTRAVSFYNVHKGKIADTPGFSSLDYEVDNAEDLNESF 227
```

```
            I+PEL L T EIS  LGRG+HTTR V   +   G +ADTPGFSSL++     E+L  +F
Sbjct: 183 AISPELGLRTNEISEHLGRGKHTTRHVELIHTSGGLVADTPGFSSLEFTDIEEEELGYTF 242

Query: 228 PELRRLSHFCKFRSCTHTHEPKCAVKEALTQGQLWQVRYDNYLQFLSEIESRRETY 283
           P++R  S  CKFR C H  EPKCAVK+A+  G+L Q RYD+Y++F++EI+ R+  Y
Sbjct: 243 PDIREKSSSCKFRGCLHLKEPKCAVKQAVEDGELKQYRYDHYVEFMTEIKDRKPRY 298
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5707> which encodes the amino acid sequence <SEQ ID 5708>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2290(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 230/290 (79%), Positives = 257/290 (88%)

Query: 1   MQGRIVKSLAGFYYVESDGVVYQTRARGNFRKKGQIPYVGDWVEFSSQDQSEGYILSIEE 60
           +QG+I+KSLAGFYYVES+G VYQTRARGNFRK+G+ PYVGD V+FS++D SEGYIL+I
Sbjct: 1   LQGKIIKSLAGFYYVESEGQVYQTRARGNFRKRGETPYVGDIVDFSAEDNSEGYILAIHP 60

Query: 61  RKNSLVRPPIVNIDQAVVIMSAKEPDFNANLLDRFLVLLEYKMIQPIIYISKLDLLDDLV 120
           RKNSLVRPPIVNIDQAVVIMSAKEP+FN+NLLDRFL+LLE+K I P++YISK+DLLD
Sbjct: 61  RKNSLVRPPIVNIDQAVVIMSAKEPEFNSNLLDRFLILLEHKAIHPVVYISKMDLLDSPE 120

Query: 121 VIDDIREHYQNIGYVFCYSQEELLPLLANKVTVFMGQTGVGKSTLLNKIAPELKLETGEI 180
            I  I   YQ IGY F  S EELLPLLA+K+TVFMGQTGVGKSTLLN+IAPEL LE GEI
Sbjct: 121 EIKAIGRQYQAIGYDFVTSLEELLPLLADKITVFMGQTGVGKSTLLNRIAPELALEIGEI 180

Query: 181 SGSLGRGRHTTRAVSFYNVHKGKIADTPGFSSLDYEVDNAEDLNESFPELRRLSHFCKFR 240
           S SLGRGRHTTRAVSFYN H GKIADTPGFSSLDY++ NAEDLNE+FPELRRLSH CKFR
Sbjct: 181 SDSLGRGRHTTRAVSFYNTHGGKIADTPGFSSLDYDIANAEDLNEAFPELRRLSHECKFR 240

Query: 241 SCTHTHEPKCAVKEALTQGQLWQVRYDNYLQFLSEIESRRETYKKVIKRK 290
           SCTHTHEPKCAVK AL  G+LW VRY++YLQFLSEIE+RRETYKKVIKRK
Sbjct: 241 SCTHTHEPKCAVKAALETGELWPVRYEHYLQFLSEIENRRETYKKVIKRK 290
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1838

A DNA sequence (GBSx1945) was identified in *S.agalactiae* <SEQ ID 5709> which encodes the amino acid sequence <SEQ ID 5710>. This protein is predicted to be rRNA. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.43   Transmembrane  259 - 275 ( 259 - 275)

----- Final Results -----
               bacterial membrane --- Certainty=0.1171(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15937 GB:Z99124 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 95/278 (34%), Positives = 147/278 (52%), Gaps = 16/278 (5%)
```

```
Query: 14  SYFACPKCQNPLIKESN-SLKCSDN-HCFDLSKFGYVNLLGGKKVDEHYDKKSFENR-QL 70
           S F CP C + +   S  SL C++  H FDLS+ GYVN L  K V   Y  + FE R +L
Sbjct: 8   SMFRCPLCDSSMDAASGKSLICTERGHTFDLSRHGYVNFLT-KPVKTSYGAELFEARSRL 66

Query: 71  VLENGYYNHILEAISKVLENNSQFH---SVLDIGCGEGFYSRQLVNKHEKTFLAF----D 123
           + E G+++ + +AI++++ +    H   ++LD GCGEG +   L         A     D
Sbjct: 67  IGECGFFDPLHDAIAELISHPKSGHEAFTILDSGCGEGSHLNALCGFDYAGKAAIGTGID 126

Query: 124 ISKDSIQLAAKSDQSRLVKWFVSDLANLPIQDSSIDIILDIFSPANYKEFRRVLSDDGIL 183
           +SKD I  A+K+ +  +  W V+D+A  P  D   D++L IFSP+NY EF R+L +DG+L
Sbjct: 127 LSKDGILKASKAFKDLM--WAVADVARAPFHDRQFDVVLSIFSPSNYAEFHRLLKNDGML 184

Query: 184 VKVVPVAEHVQELREKASQYLKQKDYSNQKILDHFRENFEIISEQKVVQSYNCSQQERQA 243
           +KVVP ++++ ELR+       ++ YSN   ++ F  N       ++       QQ
Sbjct: 185 IKVVPRSDYLIELRQFLYTDSPRRTYSNTAAVERFTANAAHSRPVRLRYVKTLDQQAIHW 244

Query: 244 FIDMTPLLFSVDKTTIDW---ASISEITVGALIVIGKK 278
            + MTPL +S  K  +        ++ITV   I+IG K
Sbjct: 245 LLKMTPLAWSAPKDRVSLLKEMKSADITVDVDILIGMK 282
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1839

A DNA sequence (GBSx1946) was identified in *S.agalactiae* <SEQ ID 5711> which encodes the amino acid sequence <SEQ ID 5712>. This protein is predicted to be dimethyladenosine transferase (ksgA). Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3257(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB11818 GB:Z99104 dimethyladenosine transferase [Bacillus subtilis]
 Identities = 157/284 (55%), Positives = 215/284 (75%), Gaps = 2/284 (0%)

Query: 3   IADKTVTRAILERHGFTFKKSFGQNFLTDTNILQKIVDTAEIDKGVNVIEIGPGIGALTE 62
           IA    T+ IL+++GF+FKKS GQNFL DTNIL +IVD AE+ +   VIEIGPGIGALTE
Sbjct: 5   IATPIRTKEILKKYGFSFKKSLGQNFLIDTNILNRIVDHAEVTEKTGVIEIGPGIGALTE 64

Query: 63  FLAENAAEVMAFEIDDRLIPILADTLARFDNVQVVNQDILKADLQTQIQA-FKNPDLPIK 121
            LA+ A +V+AFEID RL+PIL DTL+ ++NV V++QD+LKAD+++ I+  F++ D  I
Sbjct: 65  QLAKRAKKVVAFEIDQRLLPILKDTLSPYENVTVIHQDVLKADVKSVIEEQFQDCD-EIM 123

Query: 122 VVANLPYYITTPILMHLIESKIPFAEFVVMIQKEVADRISAMPNTKAYGSLSIAVQYYMT 181
           VVANLPYY+TTPI+M L+E  +P    VVM+QKEVA+R++A P++K YGSLSIAVQ+Y
Sbjct: 124 VVANLPYYVTTPIIMKLLEEHLPLKGIVVMLQKEVAERMAADPSSKEYGSLSIAVQFYTE 183

Query: 182 AKVSFIVPRTVFVPAPNVDSAILKMVRRDQPVVSVQDEDFFFRVSKVAFVHRRKTLWNNL 241
           AK   IVP+TVFVP PNVDSA+++++ RD P V V++E FFF++ K +F  RRKTL NNL
Sbjct: 184 AKTVMIVPKTVFVPQPNVDSAVIRLILRDGPAVDVENESFFFQLIKASFAQRRKTLLNNL 243

Query: 242 TSHFGKSEDTKAKLEKALEIAKIKPSIRGEALSIPDFASLADAL 285
            ++  + +  K+ +E+ LE   I    RGE+LSI +FA+L++ L
Sbjct: 244 VNNLPEGKAQKSTIEQVLEETNIDGKRRGESLSIEEFAALSNGL 287
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5713> which encodes the amino acid sequence <SEQ ID 5714>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2420(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 257/290 (88%), Positives = 275/290 (94%)

Query: 1   MRIADKTVTRAILERHGFTFKKSFGQNFLTDTNILQKIVDTAEIDKGVNVIEIGPGIGAL 60
            MRIAD +VT+A+L+RHGFTFKKSFGQNFLTDTNILQKIVDTAEID+ VNVIEIGPGIGAL
Sbjct: 9   MRIADYSVTKAVLDRHGFTFKKSFGQNFLTDTNILQKIVDTAEIDQNVNVIEIGPGIGAL 68

Query: 61  TEFLAENAAEVMAFEIDDRLIPILADTLARFDNVQVVNQDILKADLQTQIQAFKNPDLPI 120
            TEFLAENAAEVMAFEIDDRL+PILADTL  FDNVQVVNQDILKADLQTQI+ FKNPDLPI
Sbjct: 69  TEFLAENAAEVMAFEIDDRLVPILADTLRDFDNVQVVNQDILKADLQTQIKQFKNPDLPI 128

Query: 121 KVVANLPYYITTPILMHLIESKIPFAEFVVMIQKEVADRISAMPNTKAYGSLSIAVQYYM 180
            KVVANLPYYITTPILMHLIESKIPF EFVVM+Q+EVADRISA PNTKAYGSLSIAVQYYM
Sbjct: 129 KVVANLPYYITTPILMHLIESKIPFQEFVVMMQREVADRISAEPNTKAYGSLSIAVQYYM 188

Query: 181 TAKVSFIVPRTVFVPAPNVDSAILKMVRRDQPVVSVQDEDFFFRVSKVAFVHRRKTLWNN 240
            TAKV+FIVPRTVFVPAPNVDSAILKMVRRDQP++ V+DEDFFFRVS+++FVHRRKTLWNN
Sbjct: 189 TAKVAFIVPRTVFVPAPNVDSAILKMVRRDQPLIEVKDEDFFFRVSRLSFVHRRKTLWNN 248

Query: 241 LTSHFGKSEDTKAKLEKALEIAKIKPSIRGEALSIPDFASLADALKEVGI 290
            LTSHFGKSED KAKLEK L +A IKPSIRGEALSI DF  LADALKEVG+
Sbjct: 249 LTSHFGKSEDIKAKLEKGLALADIKPSIRGEALSIQDFGKLADALKEVGL 298
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1840

A DNA sequence (GBSx1947) was identified in *S.agalactiae* <SEQ ID 5715> which encodes the amino acid sequence <SEQ ID 5716>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0736(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1841

A DNA sequence (GBSx1948) was identified in *S.agalactiae* <SEQ ID 5717> which encodes the amino acid sequence <SEQ ID 5718>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3031(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB11817 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 81/179 (45%), Positives = 117/179 (65%), Gaps = 4/179 (2%)

Query: 7   IQEVIVVEGKDDTANLRRFYNVDTYETRGSAIDEDDLERIERLHNLRGVIVFTDPDYNGE 66
            I+E+IVVEG+DDTA ++   + DT ET GSAID+  +++I      RGVI+ TDPD+ GE
Sbjct: 3   IKEIIVVEGRDDTARIKLAVDADTIETNGSAIDDHVIDQIRLAQKTRGVIILTDPDFPGE 62

Query: 67  RIRKIIMNAIPTVRHAFLNRDEAKPGSKTKGRSLGVEHASFEDLQKALSKVTQHFDDEDH 126
            +IRK I  A+P  +HAFL +  AKP +K   R +GVEHAS E ++  L  V +  + +
Sbjct: 63  KIRKTISEAVPGCKHAFLPKHLAKPKNK---RGIGVEHASVESIRACLENVHEEMEAQPS 119

Query: 127 FDITQADLIRWGFITASDSRKRREYLGNQLRIGYSNGKQLLKRLRLFGVTKAEVEECME 185
             DI+  DLI  G I    ++ RRE LG+ L+IGY+NGKQL KRL++F + K++    ++
Sbjct: 120 -DISAEDLIHAGLIGGPAAKCRRERLGDLLKIGYTNGKQLQKRLQMFQIKKSDFMSALD 177
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5719> which encodes the amino acid sequence <SEQ ID 5720>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1474(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 146/187 (78%), Positives = 165/187 (88%)

Query: 1   MMKKIDIQEVIVVEGKDDTANLRRFYNVDTYETRGSAIDEDDLERIERLHNLRGVIVFTD 60
            + +KI+IQEV+VVEGKDDTANLRRFY VDTYETRGSAI E+DLERI RL++LRGVIV TD
Sbjct: 15  LTEKINIQEVLVVEGKDDTANLRRFYEVDTYETRGSAITEEDLERINRLNDLRGVIVLTD 74

Query: 61  PDYNGERIRKIIMNAIPTVRHAFLNRDEAKPGSKTKGRSLGVEHASFEDLQKALSKVTQH 120
            PDYNGERIRK+IM A+PT RHAFLNR+EA P SK+KGRSLGVEHA+FEDLQKAL+ VTQ
Sbjct: 75  PDYNGERIRKLIMAAVPTARHAFLNRNEAVPSSKSKGRSLGVEHANFEDLQKALAHVTQQ 134

Query: 121 FDDEDHFDITQADLIRWGFITASDSRKRREYLGNQLRIGYSNGKQLLKRLRLFGVTKAEV 180
            +DDE +FDI Q DLIR G + ASDSRKRREYLG +LRIGY+NGKQLLKRL LFG+T AEV
Sbjct: 135 YDDESYFDIRQTDLIRLGLLMASDSRKRREYLGEKLRIGYANGKQLLKRLELFGITLAEV 194

Query: 181 EECMEGY 187
            EE ME Y
Sbjct: 195 EEVMETY 201
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1842

A DNA sequence (GBSx1949) was identified in *S.agalactiae* <SEQ ID 5721> which encodes the amino acid sequence <SEQ ID 5722>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4955(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10139> which encodes amino acid sequence <SEQ ID 10140> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB11815 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 115/254 (45%), Positives = 172/254 (67%)

Query: 28  IFDTHTHLNVENFEGKIDEEINLASELGVTKMNVVGFDQDTISKSLELSSQYAQVYSTIG 87
            +FDTH HLN E ++  ++E I  A    V ++ VVGFD+ TI++++E+  +Y  +Y+ IG
Sbjct: 2   LFDTHAHLNAEQYDTDLEEVIERAKAEKVERIVVVGFDRPTITRAMEMIEEYDFIYAAIG 61

Query: 88  WHPTEAGSYDDNIESMIISHLENPKVIALGEIGLDYYWMEDPKDIQIEVFKRQIELSKEY 147
            WHP +A    +   + I     + KV+A+GE+GLDY+W + PKDIQ EVF+ QI L+KE
Sbjct: 62  WHPVDAIDMTEEDLAWIKELSAHEKVVAIGEMGLDYHWDKSPKDIQKEVFRNQIALAKEV 121

Query: 148 NLPFVVHTRDALEDTYEVIKESGVGPFGGIMHSFSGSLEMAQKFIDLGMMISFSGVVTFK 207
            NLP ++H RDA ED   ++KE G    GGIMH F+GS E+A++ + +   +SF G VTFK
Sbjct: 122 NLPIIIHNRDATEDVVTILKEEGAEAVGGIMHCFTGSAEVARECMKMNFYLSFGGPVTFK 181

Query: 208 KALDVQEAARELPLDKILVETDAPYLAPVPKRGRENKTAYTRYVVEKIAELRGITVEEVA 267
             A   +E  +E+P D++L+ETD P+L P P RG+ N+ +Y +YV E+IAEL+ +T EE+A
Sbjct: 182 NAKKPKEVVKEIPNDRLLIETDCPFLTPHPFRGKRNEPSYVKYVAEQIAELKEMTFEEIA 241

Query: 268 EATYQNAVRIFRLD 281
              T +NA R+FR++
Sbjct: 242 SITTENAKRLFRIN 255
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5723> which encodes the amino acid sequence <SEQ ID 5724>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2817(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 190/258 (73%), Positives = 227/258 (87%)

Query: 24  DMIKIFDTHTHLNVENFEGKIDEEINLASELGVTKMNVVGFDQDTISKSLELSSQYAQVY 83
            + + IFDTHTHLNV  F+G   EE+ LA E+GV   NVVGFDQ TIS +L L+++YA +Y
Sbjct: 38  EKLTIFDTHTHLNVAEFQGHETEELTLAQEMGVAYHNVVGFDQATISGALTLANKYANIY 97

Query: 84  STIGWHPTEAGSYDDNIESMIISHLENPKVIALGEIGLDYYWMEDPKDIQIEVFKRQIEL 143
            +TIGWHPTEAGSY + +E  I+S L + KVIALGEIGLDYYWMEDPK++QIEVFKRQ++L
Sbjct: 98  ATIGWHPTEAGSYSEAVEEAIVSQLSHSKVIALGEIGLDYYWMEDPKEVQIEVFKRQMQL 157

Query: 144 SKEYNLPFVVHTRDALEDTYEVIKESGVGPFGGIMHSFSGSLEMAQKFIDLGMMISFSGV 203
            +K+++LPFVVHTRDALEDTYEVIK +GVGP GGIMHS+SGSLEMA++FI+LGMMISFSGV
Sbjct: 158 AKDHDLPFVVHTRDALEDTYEVIKAAGVGPRGGIMHSYSGSLEMAERFIELGMMISFSGV 217

Query: 204 VTFKKALDVQEAARELPLDKILVETDAPYLAPVPKRGRENKTAYTRYVVEKIAELRGITV 263
            VTFKKALD+QEAA+ LPLDKILVETDAPYL PVPKRG++N TAYTRYVV+KIAELRG+TV
Sbjct: 218 VTFKKALDIQEAAQHLPLDKILVETDAPYLTPVPKRGKQNHTAYTRYVVDKIAELRGMTV 277
```

```
Query: 264 EEVAEATYQNAVRIFRLD 281
           EEVA+AT  NA R+F+LD
Sbjct: 278 EEVAKATTANAKRVFKLD 295
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1843

A DNA sequence (GBSx1950) was identified in *S.agalactiae* <SEQ ID 5725> which encodes the amino acid sequence <SEQ ID 5726>. This protein is predicted to be endosome-associated protein. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5142(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1844

A DNA sequence (GBSx1951) was identified in *S.agalactiae* <SEQ ID 5727> which encodes the amino acid sequence <SEQ ID 5728>. This protein is predicted to be CG17785 gene product. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4730(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1845

A DNA sequence (GBSx1952) was identified in *S.agalactiae* <SEQ ID 5729> which encodes the amino acid sequence <SEQ ID 5730>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4032(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB01041 GB:AB022220 gene_id:MLN21.14~unknown protein
          [Arabidopsis thaliana]
 Identities = 49/185 (26%), Positives = 85/185 (45%), Gaps = 46/185 (24%)

Query: 5    LTDLDRVNIAKQEYELGSQLDTLVKIMSQDKVLPIGKVAHVQ------DGGKETGEQIYT 58
            L  +D V+   + + ELGS+    + +M+        K+  V+      D  K+   Q++
Sbjct: 154 LEGIDSVDSGRVKIELGSRGLMDLCVMASKLAYENAKMNLVEFLDCWNDYQKQMSTQVFV 213

Query: 59   ITPNGTLDKPEDVKEVTVLFKGSTAPFGGDDWKTD----WFKNDIPIASKL---LLKKFG 111
             T     DK +D   + + F+G T PF  DDW TD    W+  ++P   KL    L+  G
Sbjct: 214 FT-----DKQKDANLIVISFRG-TEPFDADDWGTDFDYSWY--EVPNVGKLHMGFLEAMG 265

Query: 112 ---------------SQSVSHKQGTKQ-----LEQSAH-----LLKEVMNKYPNAKISVY 146
                            Q+ S ++ +K+     +E+SA+     +LK +++++ NA+  V
Sbjct: 266 LGNRDDTTTFHYNLFEQTSSEEENSKKNLLDMVERSAYYAVRVILKRLLSEHENARFVVT 325

Query: 147 GHSLG 151
            GHSLG
Sbjct: 326 GHSLG 330
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1846

A DNA sequence (GBSx1953) was identified in *S.agalactiae* <SEQ ID 5731> which encodes the amino acid sequence <SEQ ID 5732>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.97    Transmembrane   12 -  28 (   5 -  33)

----- Final Results -----
               bacterial membrane --- Certainty=0.4588(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10141> which encodes amino acid sequence <SEQ ID 10142> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8909> and protein <SEQ ID 8910> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 4
McG: Discrim Score:     14.01
GvH: Signal Score (-7.5): -5.55
     Possible site: 46
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 1 value:  -8.97 threshold:  0.0
   INTEGRAL    Likelihood = -8.97    Transmembrane    6 -  22 (   1 -  27)
```

```
   PERIPHERAL  Likelihood =  9.49      84
 modified ALOM score:   2.29

*** Reasoning Step: 3

----- Final Results -----
                 bacterial membrane --- Certainty=0.4588(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

SEQ ID 8910 (GBS32) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 10 (lane 2; MW 15.6kDa).

GBS32-His was purified as shown in Figure 191, lane 8.

Example 1847

A DNA sequence (GBSx1954) was identified in *S.agalactiae* <SEQ ID 5733> which encodes the amino acid sequence <SEQ ID 5734>. This protein is predicted to be extramembranal protein (dltD). Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.24   Transmembrane   12 -  28 (   4 -  31)

----- Final Results -----
                 bacterial membrane --- Certainty=0.5097(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC29041 GB:AF050517 unknown [Streptococcus mutans]
 Identities = 242/421 (57%), Positives = 309/421 (72%), Gaps = 1/421 (0%)

Query: 1   MLKRLGKVFGPLVCALLLLVGLYFVFPVSQ-PHHLGKEKNSAVALTKAGFKSRVQKVRAF 59
           MLKRL  + GP+ CAL+L+  L   +P     H+  +EKN AVAL+ + FKS  +K+RA
Sbjct: 1   MLKRLWLILGPVFCALVLVFSLIMFYPAKHLSHNYNEEKNDAVALSPSSFKSTNKKMRAL 60

Query: 60  SDPKANFVPFFGSSEWLRFDAMHPSVLAEAYNRSYIPYLLGQKGAASLTQYYGIQQIKGQ 119
           SD +  FVPFFGSSEW R D MHPSVLAE YNRSY PYLLGQKG+ SL+ Y+G+QQI  Q
Sbjct: 61  SDKRHLFVPFFGSSEWQRIDNMHPSVLAERYNRSYRPYLLGQKGSTSLSHYFGMQQIGNQ 120

Query: 120 IKNKKAIYVISPQWFVRKGANKGAFQNYFSNDQTIRFLQNQTGTTYDRYAARRLLKLYPE 179
           IKNKKA+YVISPQWFV KG +  AFQ YFS++Q   FL NQTG+T DRYAA+RLL + P
Sbjct: 121 IKNKKAVYVISPQWFVPKGTSPIAFQQYFSSEQLADFLLNQTGSTADRYAAKRLLDIKPS 180

Query: 180 ASMSDLIEKVADGQKLSNKDKQRLKFNDWVFEKTDAIFSYLPLGKTYNQAIMPHVGKLPK 239
           +++  +I+K+A G+ L++ D+  L+       +K DA+F  L     Y + ++PHV KLPK
Sbjct: 181 SNLQGMIKKIAAGKTLNSFDRASLRLIKSFLKKEDALFGSLTFSDNYERRVLPHVKKLPK 240

Query: 240 AFSYNHLSRIASQDAKVATRSNQFGIDDRFYQTRIKKHLKKLKGSQRHFNYTKSPEFNDL 299
            FSY  LS+IAS+D +  T++NQF I+D FY  RIK  LK+LKG Q+  +Y +SPE+NDL
Sbjct: 241 HFSYGTLSQIASKDGQRLTKTNQFEINDHFYNKRIKGQLKRLKGFQKQLSYLQSPEYNDL 300

Query: 300 QLVLNEFSKQNTDVLFVIPPVNKKWTDYTGLDQKMYQKSVEKIKHQLQSQGFNHIADLSR 359
           QL L + +K  T V+FVIPPVN KW +YTGL Q MYQK+VEKIK+QLQSQGF++IADLS+
Sbjct: 301 QLALTQLAKSKTKVIFVIPPVNAKWVEYTGLSQDMYQKTVEKIKYQLQSQGFDNIADLSK 360

Query: 360 DGGKPYFMQDTIHLGWNGWLELDKHINPFLTEENSKPNYHINNKFLKKSWAKYTGRPSDYK 420
           +G +PYFMQDTIHLGWNGWL  DK +NPFL+++  +P Y INN FL K WA YTG P  +K
Sbjct: 361 NGDQPYFMQDTIHLGWNGWLAFDKEVNPFLSKKQLQPAYKINNHFLSKKWATYTGNPFQFK 421
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5735> which encodes the amino acid sequence <SEQ ID 5736>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-13.06    Transmembrane     7 -  23 (   1 -  31)

----- Final Results -----
              bacterial membrane --- Certainty=0.6222(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 209/410 (50%), Positives = 278/410 (66%)

Query: 1   MLKRLGKVFGPLVCALLLLVGLYFVFPVSQPHHLGKEKNSAVALTKAGFKSRVQKVRAFS 60
           MLKRL  + GPL+ A +L+V   F FP    H + +EK +AVA+T + FK+ + K +A S
Sbjct: 1   MLKRLWLILGPLLIAFVLVVITIFSFPTQLDHSIAQEKANAVAITDSSFKNGLIKRQALS 60

Query: 61  DPKANFVPFFGSSEWLRFDAMHPSVLAEAYNRSYIPYLLGQKGAASLTQYYGIQQIKGQI 120
           D    FVPFFGSSEW R D+MHPSVLAE Y RSY P+L+G++G+ASL+ YYGIQQI  ++
Sbjct: 61  DETCRFVPFFGSSEWSRMDSMHPSVLAERYKRSYRPFLIGKRGSASLSHYYGIQQITNEM 120

Query: 121 KNKKAIYVISPQWFVRKGANKGAFQNYFSNDQTIRFLQNQTGTTYDRYAARRLLKLYPEA 180
           + KKAI+V+SPQWF  +G N  A Q Y SN Q I FL         ++AA+RLL+L P
Sbjct: 121 QKKKAIFVVSPQWFTAQGINPSAVQMYLSNTQVIEFLLKARTDKESQFAAKRLLELNPGV 180

Query: 181 SMSDLIEKVADGQKLSNKDKQRLKFNDWVFEKTDAIFSYLPLGKTYNQAIMPHVGKLPKA 240
           S S+L++KV+ G+ LS  D+  LK    V  + +++FS+L     Y + I+P V  LPK
Sbjct: 181 SKSNLLKKVSKGKSLSRLDRAILKCQHQVALREESLFSFLGKSTNYEKRILPRVKGLPKV 240

Query: 241 FSYNHLSRIASQDAKVATRSNQFGIDDRFYQTRIKKHLKKLKGSQRHFNYTKSPEFNDLQ 300
           FSY  L+ +A++  ++AT +N+FGI + FY+ RI       K  Q +++Y  SPE+ND Q
Sbjct: 241 FSYKQLNALATKRGQLATTNNRFGIKNTFYRKRIAPKYNLYKNFQVNYSYLASPEYNDFQ 300

Query: 301 LVLNEFSKQNTDVLFVIPPVNKKWTDYTGLDQKMYQKSVEKIKHQLQSQGFNHIADLSRD 360
           L+L+EF+K+ TDVLFVI PVNK W DYTGL+Q  YQ +V KIK QL+SQGF+ IAD S+D
Sbjct: 301 LLLSEFAKRKTDVLFVITPVNKAWADYTGLNQDKYQAAVRKIKFQLKSQGFHRIADFSKD 360

Query: 361 GGKPYFMQDTIHLGWNGWLELDKHINPFLTEENSKPNYHINNKFLKKSWA 410
           GG+ YFMQDTIHLGWNGWL  DK + PFL  +   PNY +N  F  K WA
Sbjct: 361 GGESYFMQDTIHLGWNGWLAFDKKVQPFLETKQPVPNYKMNPYFYSKIWA 410
```

A related GBS gene <SEQ ID 8911> and protein <SEQ ID 8912> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 6
McG: Discrim Score:      15.50
GvH: Signal Score (-7.5): -4.52
     Possible site: 31
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 1 value: -10.24 threshold:  0.0
   INTEGRAL    Likelihood =-10.24    Transmembrane    12 -  28 (   4 -  31)
   PERIPHERAL  Likelihood =  8.33      301
 modified ALOM score:   2.55

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.5097(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
57.5/76.3% over 420aa
                                             Streptococcus mutans
  GP|3403204| unknown Insert characterized
```

```
ORF00336(301 - 1560 of 1860)
GP|3403204|gb|AAC29041.1||AF050517(1 - 421 of 421) unknown {Streptococcus mutans}
%Match = 41.0
%Identity = 57.5  %Similarity = 76.2
Matches = 242  Mismatches = 99  Conservative Sub.s = 79

33        63        93        123       153       183       213       243
FSGFLDLLWFPQPHNK**GVL*WILNQKY*QLLMTYLWRMFLL*WMKTYLTQEF*TAWVLLN*LLSWKATLILIFRLRNL

273       303       333       363                 420       450       480
VVMTGTQLIKLLLE*RSSAMLKRLGKVFGPLVCALLLLVGLYFVFPVSQ-PHHLGKEKNSAVALTKAGFKSRVQKVRAFS
                   |||||  ::||: |||:|:  |   :|  :  |:  :||| ||||: : |||  :|:||:|
                   MLKRLWLILGPVFCALVLVFSLIMFYPAKHLSHNYNEEKNDAVALSPSSFKSTNKKMRALS
                           10        20        30        40        50        60

510       540       570       600       630       660       690       720
DPKANFVPFFGSSEWLRFDAMHPSVLAEAYNRSYIPYLLGQKGAASLTQYYGIQQIKGQIKNKKAIYVISPQWFVRKGAN
| :  |||||||||| | | |||||||| ||||| ||||||||: ||::|:|:|||  |||||||:||||||||| || :
DKRHLFVPFFGSSEWQRIDNMHPSVLAERYNRSYRPYLLGQKGSTSLSHYFGMQQIGNQIKNKKAVYVISPQWFVPKGTS
                 80        90       100       110       120       130       140

750       780       810       840       870       900       930       960
KGAFQNYFSNDQTIRFLQNQTGTTYDRYAARRLLKLYPEASMSDLIEKVADGQKLSNKDKQRLKFNDWVFEKTDAIFSYL
  ||| |||::|   || ||||:| |||||:||| : | :::  :|:|:| |: |:: |:  |::    ::| ||:|  |
PIAFQQYFSSEQLADFLLNQTGSTADRYAAKRLLDIKPSSNLQGMIKKIAAGKTLNSFDRASLRLIKSFLKKEDALFGSL
                160       170       180       190       200       210       220

990       1020      1050      1080      1110      1140      1170      1200
PLGKTYNQAIMPHVGKLPKAFSYNHLSRIASQDAKVATRSNQFGIDDRFYQTRIKKHLKKLKGSQRHFNYTKSPEFNDLQ
 :   | : ::||| |||| |||  ||:|||:| :  |::||| |:| ||  ||| :||:||| |::::| :|||:||||
TFSDNYERRVLPHVKKLPKHFSYGTLSQIASKDGQRLTKTNQFEINDHFYNKRIKGQLKRLKGFQKQLSYLQSPEYNDLQ
                240       250       260       270       280       290       300

1230      1260      1290      1320      1350      1380      1410      1440
LVLNEFSKQNTDVLFVIPPVNKKWTDYTGLDQKMYQKSVEKIKHQLQSQGFNHIADLSRDGGKPYFMQDTIHLGWNGWLE
| | :::|  | |:||||||| || :|||| | ||||:|||||:|||||||::|||||::| :||||||||||||||||
LALTQLAKSKTKVIFVIPPVNAKWVEYTGLSQDMYQKTVEKIKYQLQSQGFDNIADLSKNGDQPYFMQDTIHLGWNGWLA
                320       330       340       350       360       370       380

1470      1500      1530      1560      1590      1620      1650      1680
LDKHINPFLTEENSKPNYHINNKFLKKSWAKYTGRPSDYK*IVESDDL*H*SY*SSFLISLYLVILR*LIHVL*FFIYNE
:|| :||||:::  :| | ||| || | || ||| |  :|
FDKEVNPFLSKKQLQPAYKINNHFLSKKWATYTGNPFQFK
                400       410       420
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1848

A DNA sequence (GBSx1955) was identified in *S.agalactiae* <SEQ ID 5737> which encodes the amino acid sequence <SEQ ID 5738>. This protein is predicted to be d-alanyl carrier protein (dltC). Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1061(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC05776 GB:AF051356 D-alanyl carrier protein [Streptococcus mutans]
```

```
 Identities = 65/79 (82%), Positives = 74/79 (93%)

Query: 1  MDIKSEVLAIIDDLFMEDVSSMMDEDLFDAGVLDSMGTVELIVELESHFNIDIPIAEFGR 60
           MDIKSEVL IID+LFMEDVS MMDEDLFDAGVLDSMGTVELIVELE+HF+I +P++EFGR
Sbjct: 1  MDIKSEVLKIIDELFMEDVSDMMDEDLFDAGVLDSMGTVELIVELENHFDITVPVSEFGR 60

Query: 61 NDWNTANKIVAGVTELCNA 79
           +DWNTANKI+ G+TEL NA
Sbjct: 61 DDWNTANKIIEGITELRNA 79
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5739> which encodes the amino acid sequence <SEQ ID 5740>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3976(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 57/79 (72%), Positives = 65/79 (82%)

Query: 1  MDIKSEVLAIIDDLFMEDVSSMMDEDLFDAGVLDSMGTVELIVELESHFNIDIPIAEFGR 60
           M I+  V+ + D LFMEDVS MMDEDLFDAGVLDS+GTVELIVELES FNI +PI+EFGR
Sbjct: 1  MSIEETVIELFDRLFMEDVSEMMDEDLFDAGVLDSLGTVELIVELESTFNIKVPISEFGR 60

Query: 61 NDWNTANKIVAGVTELCNA 79
           +DWNT  KIV GV EL +A
Sbjct: 61 DDWNTVTKIVQGVEELQHA 79
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1849

A DNA sequence (GBSx1956) was identified in *S.agalactiae* <SEQ ID 5741> which encodes the amino acid sequence <SEQ ID 5742>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.55   Transmembrane   93 - 109 (  91 - 117)
   INTEGRAL    Likelihood = -7.64   Transmembrane   21 -  37 (  19 -  39)
   INTEGRAL    Likelihood = -6.79   Transmembrane  390 - 406 ( 387 - 410)
   INTEGRAL    Likelihood = -5.20   Transmembrane   41 -  57 (  40 -  59)
   INTEGRAL    Likelihood = -2.07   Transmembrane  203 - 219 ( 200 - 221)
   INTEGRAL    Likelihood = -1.65   Transmembrane   65 -  81 (  65 -  81)
   INTEGRAL    Likelihood = -0.75   Transmembrane  125 - 141 ( 125 - 141)

----- Final Results -----
               bacterial membrane --- Certainty=0.4418(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5743> which encodes the amino acid sequence <SEQ ID 5744>. Analysis of this protein sequence reveals the following:

```
     Possible site: 57
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.14   Transmembrane  387 - 403 ( 382 - 409)
   INTEGRAL    Likelihood = -9.66   Transmembrane   18 -  34 (  15 -  37)
   INTEGRAL    Likelihood = -5.95   Transmembrane   64 -  80 (  63 -  81)
```

```
  INTEGRAL    Likelihood = -5.63   Transmembrane   92 - 108 (  89 - 114)
  INTEGRAL    Likelihood = -1.97   Transmembrane   40 -  56 (  40 -  56)

----- Final Results -----
              bacterial membrane --- Certainty=0.5055(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC05775 GB:AF051356 integral membrane protein [Streptococcus mutans]
 Identities = 246/413 (59%), Positives = 319/413 (76%)

Query: 1   MMMFFSHIPYMEPYGNPIYFVYLILAFLPVIIGIFKQKRLSTYETLVSLVFILFMFGGDH 60
           M+ FF ++P++E YGNP YF Y+ILA LP+ IG+F +KR   YE  VSL+FI+ M  G+
Sbjct: 1   MIDFFKNLPHLEAYGNPQYFFYIILAVLPIFIGLFFKKRFPLYEAFVSLIFIVLMLTGEK 60

Query: 61  YQQLVAFLFYLLWQIISVFAYQKYRENANSAGVFYLAIAMALFPLIWVKVAPLTGPSSQT 120
             Q+ A  FY++WQI  V++Y+ YR++ ++  +FYL + M++ PL  VK+ P    + Q+
Sbjct: 61  SHQIFALFFYIIWQIFCVYSYKFYRKSRDNKWIFYLHVFMSILPLSLVKITPAIWTNQQS 120

Query: 121 LFSFLGISYLTFKSIGMIIEMRDGTLQEVRLPDFIRFMIFFPTFSSGPIDRFRHFQEDYH 180
           LF FLGISYLTF+S+GMI+EMRDG L      +FIRFM+F PTFSSGPIDRFR F +DY
Sbjct: 121 LFGFLGISYLTFRSVGMIMEMRDGVLTSFTFWEFIRFMLFMPTFSSGPIDRFRRFNDDYE 180

Query: 181 KLPERDDYFAMLNKAVMYLMLGFLYKHIISYCLGGILLPLLENKALMVGGYFNKETILVM 240
           K+P++D+   ML ++V Y+MLGF YK +++  LG ++LP L+  AL  GG+FN  T+ VM
Sbjct: 181 KIPDKDELLDMLEQSVHYIMLGFFYKFVLAQILGTMILPGLKEMALQKGGWFNWPTLGVM 240

Query: 241 YVYGLNLFFDFAGYSMFAIGISYLLGIRTPENFNMPFLSASLKDFWNRWHMSLSFWFRDY 300
           YVYGL+LFFDFAGYSMFAI IS  +GI++P NFN PF S  LK+FWNRWHMSLSFWFRD+
Sbjct: 241 YVYGLDLFFDFAGYSMFAIAISNFMGIKSPTNFNQPFKSQDLKEFWNRWHMSLSFWFRDF 300

Query: 301 VFMRLVHLLIKHKTFKNRNVTSGVAYLVNMLVMGFWHGLTWYYIAYGLFHGIGLIINDAW 360
           VFMRLV +L+K+K FKNRNVTS VAY+VNML+MGFWHG+TWYYI YGLFHG+GL++NDAW
Sbjct: 301 VFMRLVKVLVKNKVFKNRNVTSSVAYIVNMLIMGFWHGVTWYYITYGLFHGVGLVLNDAW 360

Query: 361 IRKKKEINRHRKKKGLSPLFQSRAFHVLCIVVTFHVVMFSLLLFSGFLNDLWF 413
           +RKKK +N+ RK K LSPL ++     L IV+TF+VVM S L+FSGFLNDLWF
Sbjct: 361 LRKKKRLNKERKAKNLSPLPENGWTRALGIVITFNVVMLSFLIFSGFLNDLWF 413
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 240/416 (57%), Positives = 317/416 (75%), Gaps = 5/416 (1%)

Query: 5   FLEKLPHLDVYGNPQYFFYLILAVLPIYIGLFFKKRFALYEIIFSLSFIVMMLTGSTFNQ 64
           F   +P+++ YGNP YF YLILA LP+ IG+F +KR + YE + SL FI+ M  G  + Q
Sbjct: 4   FFSHIPYMEPYGNPIYFVYLILAFLPVIIGIFKQKRLSTYETLVSLVFILFMFGGDHYQQ 63

Query: 65  LKSLLAYVVGQSLLVFIYKAYRKRFNHTLVFYVTVCLSIFPLFLVKLIPAISEDGHQSLF 124
           L + L Y++ Q + VF Y+ YR+  N    VFY+ + +++FPL  VK+ P ++    Q+LF
Sbjct: 64  LVAFLFYLLWQIISVFAYQKYRENANSAGVFYLAIAMALFPLIWVKVAP-LTGPSSQTLF 122

Query: 125 GFLGISYLTFRAVAMIIEMRDGVLKEFTLWEFLRFLLFFPTFSSGPIDRFKRFNEDYINI 184
            FLGISYLTF+++ MIIEMRDG L+E  L +F+RF++FFPTFSSGPIDRF+ F EDY  +
Sbjct: 123 SFLGISYLTFKSIGMIIEMRDGTLQEVRLPDFIRFMIFFPTFSSGPIDRFRHFQEDYHKL 182

Query: 185 PDRNELLDMLGQAIHYLMLGFLYKFILAYIFGSLIMPPLKELALEQGGVFNWPTLGVMYA 244
           P+R++   ML +A+ YLMLGFLYK I++Y  G +++P L+  AL  GG FN  T+ VMY
Sbjct: 183 PERDDYFAMLNKAVMYLMLGFLYKHIISYCLGGILLPLLENKALMVGGYFNKETILVMYV 242

Query: 245 FGFDLFFDFAGYTMFALAISNLMGIKSPINFDKPFKSRDLKEFWNRWHMSLSFWFRDFVF 304
           +G +LFFDFAGY+MFA+ IS L+GI++P NF+ PF S  LK+FWNRWHMSLSFWFRD+VF
Sbjct: 243 YGLNLFFDFAGYSMFAIGISYLLGIRTPENFNMPFLSASLKDFWNRWHMSLSFWFRDYVF 302

Query: 305 MRLVKLLVKNKVFKNRNVTSSVAYIINMLLMGFWHGLTWYYIAYGLFHGIGLVINDAWVR 364
           MRLV LL+K+K FKNRNVTS VAY++NML+MGFWHGLTWYYIAYGLFHGIGL+INDAW+R
Sbjct: 303 MRLVHLLIKHKTFKNRNVTSGVAYLVNMLVMGFWHGLTWYYIAYGLFHGIGLIINDAWIR 362
```

```
Query: 365 KKKNINKERRLAKKPLLP--ENKWTYALGVFITFNVVMFSFLIFSGFLDLLWFPQP 418
           KKK IN+ R+  KK L P  +++   + L + +TF+VVMFS L+FSGFL+ LWF +P
Sbjct: 363 KKKEINRHRK--KKGLSPLFQSRAFHVLCIVVTFHVVMFSLLLFSGFLNDLWFNRP 416
```

A related GBS gene <SEQ ID 8913> and protein <SEQ ID 8914> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:      3.22
GvH: Signal Score (-7.5): -4.56
     Possible site: 16
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 7 value:  -8.55 threshold:  0.0
   INTEGRAL    Likelihood = -8.55   Transmembrane   93 - 109 (  91 - 117)
   INTEGRAL    Likelihood = -7.64   Transmembrane   21 -  37 (  19 -  39)
   INTEGRAL    Likelihood = -6.79   Transmembrane  390 - 406 ( 387 - 410)
   INTEGRAL    Likelihood = -5.20   Transmembrane   41 -  57 (  40 -  59)
   INTEGRAL    Likelihood = -2.07   Transmembrane  203 - 219 ( 200 - 221)
   INTEGRAL    Likelihood = -1.65   Transmembrane   65 -  81 (  65 -  81)
   INTEGRAL    Likelihood = -0.75   Transmembrane  125 - 141 ( 125 - 141)
   PERIPHERAL  Likelihood =  1.01     322
 modified ALOM score:   2.21

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4418(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01206(313 - 1563 of 1863)
GP|2952530|gb|AAC05775.1||AF051356(4 - 419 of 420) integral membrane protein {Streptococcus
mutans}
%Match = 50.3
%Identity = 71.0  %Similarity = 86.6
Matches = 296  Mismatches = 55  Conservative Sub.s = 65

273       303       333       363       393       423       453       483
TFDTKWEN*YQRSYERGKQVIQAFLEKLPHLDVYGNPQYFFYLILAVLPIYIGLFFKKRFALYEIIFSLSFIVMMLTGST
                       |:: ||||: |||||||||:|||||||:||||||||| |||   || |||:||||  
                    MIDFFKNLPHLEAYGNPQYFFYIILAVLPIFIGLFFKKRFPLYEAFVSLIFIVLMLTGEK
                            10        20        30        40        50        60

513       543       573       603       633       663       693       723
FNQLKSLLAYVVGQSLLVFIYKAYRKRFNHTLVFYVTVCLSIFPLFLVKLIPAISEDGHQSLFGFLGISYLTFRAVAMII
 :|: :|: |:: | : |: || |||  ::  :||: | :||:|| |||: |||  : :|||||||||||||||:| ||:
SHQIFALFFYIIWQIFCVYSYKFYRKSRDNKWIFYLHVFMSILPLSLVKITPAIWTN-QQSLFGFLGISYLTFRSVGMIM
        70        80        90       100       110        120       130

753       783       813       843       873       903       933       963
EMRDGVLKEFTLWEFLRFLLFFPTFSSGPIDRFKRFNEDYINIPDRNELLDMLGQAIHYLMLGFLYKFILAYIFGSLIMP
|||||||  ||:|||:||:|| |||||||||||:|||:||  |||::|||||| |::||:||||:|||:|| |:|::|:|
EMRDGVLTSFTFWEFIRFMLFMPTFSSGPIDRFRRFNDDYEKIPDKDELLDMLEQSVHYIMLGFFYKFVLAQILGTMILP
        150       160       170       180       190       200       210

993      1023      1053      1083      1113      1143      1173      1203
PLKELALEQGGVFNWPTLGVMYAFGFDLFFDFAGYTMFALAISNLMGIKSPINFDKPFKSRDLKEFWNRWHMSLSFWFRD
 |||:||::||| ||||||||||  :|:|||||||||:|||:||||:|||||| ||::||||:|||||||||||||||||||
GLKEMALQKGGWFNWPTLGVMYVYGLDLFFDFAGYSMFAIAISNFMGIKSPTNFNQPFKSQDLKEFWNRWHMSLSFWFRD
        230       240       250       260       270       280       290

1233      1263      1293      1323      1353      1383      1413      1443
FVFMRLVKLLVKNKVFKNRNVTSSVAYIINMLLMGFWHGLTWYYIAYGLFHGIGLVINDAWVRKKKNINKERRLAKKPLL
||||||||:|||||||||||||||||||:|||:||||||:||||| ||||||:|||:||||:|||| :||||:      |
```

```
FVFMRLVKVLVKNKVFKNRNVTSSVAYIVNMLIMGFWHGVTWYYITYGLFHGVGLVLNDAWLRKKKRLNKERKAKNLSPL
        310       320       330       340       350       360       370

1473      1503      1533      1563      1593      1623      1653      1683
PENKWTYALGVFITFNVVMFSFLIFSGFLDLLWFPQPHNK**GVL*WILNQKY*QLLMTYLWRMFLL*WMKTYLTQEF*T
||| || |||: |||||||:|||||||||: |||    :|
PENGWTRALGIVITFNVVMLSFLIFSGFLNDLWFADQLSKK
        390       400       410       420
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1850

A DNA sequence (GBSx1957) was identified in *S.agalactiae* <SEQ ID 5745> which encodes the amino acid sequence <SEQ ID 5746>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2611(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10143> which encodes amino acid sequence <SEQ ID 10144> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC05774 GB:AF051356 D-alanine-D-alanyl carrier protein ligase
          [Streptococcus mutans]
 Identities = 404/510 (79%), Positives = 465/510 (90%)

Query: 5   IHDMIKTIEHFAETQADFPVYDILGEVHTYGQLKVDSDSLAAHIDSLGLVEKSPVLVFGG 64
           I DMI TIE+FA+ QA+FPVY+ILGE+HTYG+LK DSDSLAAH+D L L  KSPV+VFGG
Sbjct: 6   IKDMIATIENFAQEQAEFPVYNILGEIHTYGELKADSDSLAAHLDQLDLTAKSPVVVFGG 65

Query: 65  QEYEMLATFVALTKSGHAYIPVDQHSALDRIQAIMTVAQPSLIISIGEFPLEVDNVPILD 124
           QEY MLA+FVALTKSGHAYIP+D HSAL+RI+AI+ VA+PSL+I++ +FP++   VP++
Sbjct: 66  QEYAMLASFVALTKSGHAYIPIDHHSALERIEAILEVAEPSLVIAVDDFPIDNLQVPVIQ 125

Query: 125 VSQVSAIFEEKTPYEVTHSVKGDDNYYIIFTSGTTGLPKGVQISHDNLLSFTNWMISDDE 184
            SQ+  IF++K  Y++ H+VKGDD YYIIFTSGTTG PKGVQISHDNLLSFTNWMI+ +
Sbjct: 126 YSQLEEIFKQKLSYQINHAVKGDDTYYIIFTSGTTGKPKGVQISHDNLLSFTNWMINAEA 185

Query: 185 FSVPERPQMLAQPPYSFDLSVMYWAPTLAMGGTLFALPKTVVNDFKKLFATINELPIQVW 244
           F+ P RPQMLAQPPYSFDLSVMYWAPTLA+GGTLFALPK +  DFK+LF TIN+LPI VW
Sbjct: 186 FATPHRPQMLAQPPYSFDLSVMYWAPTLALGGTLFALPKEITADFKQLFTTINQLPIGVW 245

Query: 245 TSTPSFADMALLSNDFNSETLPQLTHFYFDGEELTVKTAQKLRQRFPKARIVNAYGPTEA 304
           TSTPSF DMA+LS+DFN++ LP LTHFYFDGEELTVKTA+KLRQRFP+ARIVNAYGPTEA
Sbjct: 246 TSTPSFVDMAMLSDDFNAQQLPHLTHFYFDGEELTVKTAKKLRQRFPQARIVNAYGPTEA 305

Query: 305 TVALSAVAITDEMLETCKRLPIGYTKDDSPTYVIDEEGHKLPNGEQGEIIIAGPAVSKGY 364
           TVALSA+A+TD+MLETCKRLPIGYTK DSPT++IDE GHKL NG+QGEII++GPAVSKGY
Sbjct: 306 TVALSALAVTDKMLETCKRLPIGYTKPDSPTFIIDESGHKLANGQQGEIIVSGPAVSKGY 365

Query: 365 LNNPEKTAEAFFQFEGLPAYHTGDLGSMTDEGLLLYGGRMDFQIKFNGYRIELEDVSQNL 424
           LNNPE+TA AFF+FEGLPAYHTGDLGSMTDEGLLLYGGRMDFQIKFNGYRIELE+VSQNL
Sbjct: 366 LNNPERTAAAFFEFEGLPAYHTGDLGSMTDEGLLLYGGRMDFQIKFNGYRIELEEVSQNL 425

Query: 425 NKSQYVKSAVAVPRYNKDHKVQNLLAYIVLKEGVRDDFERDLDLTKAIKEDLKDIMMDYM 484
           NKSQY+ SAVAVPRYNKDHKVQNLLAY+VLK+GV + FER LD+TKAIK DL+D+MMDYM
Sbjct: 426 NKSQYIASAVAVPRYNKDHKVQNLLAYVVLKDGVEEQFERALDITKAIKADLQDVMMDYM 485
```

```
Query: 485 MPSKFIYREDLPLTPNGKIDIKGLMSEVNK 514
           MPSKF+YR+DLPLTPNGKIDIKGLMSEVNK
Sbjct: 486 MPSKFLYRKDLPLTPNGKIDIKGLMSEVNK 515
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5747> which encodes the amino acid sequence <SEQ ID 5748>. Analysis of this protein sequence reveals the following:

```
        Possible site: 60
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.28   Transmembrane   92 - 108 (  91 - 108)
   INTEGRAL    Likelihood = -0.85   Transmembrane   43 -  59 (  41 -  59)

----- Final Results -----
               bacterial membrane --- Certainty=0.1914(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC05774 GB:AF051356 D-alanine-D-alanyl carrier protein ligase
           [Streptococcus mutans]
 Identities = 365/511 (71%), Positives = 438/511 (85%)

Query: 2   IKDMIDSIEQFAQTQADFPVYDCLGERRTYGQLKRDSDSIAAFIDSLALLAKSPVLVFGA 61
            IKDMI +IE FAQ QA+FPVY+ LGE  TYG+LK DSDS+AA +D L L AKSPV+VFG
Sbjct: 6   IKDMIATIENFAQEQAEFPVYNILGEIHTYGELKADSDSLAAHLDQLDLTAKSPVVVFGG 65

Query: 62  QTYDMLATFVALTKSGHAYIPVDVHSAPERILAIIEIAKPSLIIAIEEFPLTIEGISLVS 121
            Q Y MLA+FVALTKSGHAYIP+D HSA ERI AI+E+A+PSL+IA+++FP+    + ++
Sbjct: 66  QEYAMLASFVALTKSGHAYIPIDHHSALERIEAILEVAEPSLVIAVDDFPIDNLQVPVIQ 125

Query: 122 LSEIESAKLAEMPYERTHSVKGDDNYYIIFTSGTTGQPKGVQISHDNLLSFTNWMIEDAA 181
             S++E     ++ Y+  H+VKGDD YYIIFTSGTTG+PKGVQISHDNLLSFTNWMI   A
Sbjct: 126 YSQLEEIFKQKLSYQINHAVKGDDTYYIIFTSGTTGKPKGVQISHDNLLSFTNWMINAEA 185

Query: 182 FDVPKQPQMLAQPPYSFDLSVMYWAPTLALGGTLFALPKELVADFKQLFTTIAQLPVGIW 241
            F  P +PQMLAQPPYSFDLSVMYWAPTLALGGTLFALPKE+ ADFKQLFTTI QLP+G+W
Sbjct: 186 FATPHRPQMLAQPPYSFDLSVMYWAPTLALGGTLFALPKEITADFKQLFTTINQLPIGVW 245

Query: 242 TSTPSFADMAMLSDDFCQAKMPALTHFYFDGEELTVSTARKLFERFPSAKIINAYGPTEA 301
            TSTPSF DMAMLSDDF    ++P LTHFYFDGEELTV TA+KL +RFP A+I+NAYGPTEA
Sbjct: 246 TSTPSFVDMAMLSDDFNAQQLPHLTHFYFDGEELTVKTAKKLRQRFPQARIVNAYGPTEA 305

Query: 302 TVALSAIEITREMVDNYTRLPIGYPKPDSPTYIIDEDGKELSSGEQGEIIVTGPAVSKGY 361
            TVALSA+ +T +M++    RLPIGY KPDSPT+IIDE G +L++G+QGEIIV+GPAVSKGY
Sbjct: 306 TVALSALAVTDKMLETCKRLPIGYTKPDSPTFIIDESGHKLANGQQGEIIVSGPAVSKGY 365

Query: 362 LNNPEKTAEAFFTFKGQPAYHTGDIGSLTEDNILLYGGRLDFQIKYAGYRIELEDVSQQL 421
            LNNPE+TA AFF F+G PAYHTGD+GS+T++ +LLYGGR+DFQIK+ GYRIELE+VSQ L
Sbjct: 366 LNNPERTAAAFFEFEGLPAYHTGDLGSMTDEGLLLYGGRMDFQIKFNGYRIELEEVSQNL 425

Query: 422 NQSPMVASAVAVPRYNKEHKVQNLLAYIVVKDGVKERFDRELELTKAIKASVKDHMMSYM 481
            N+S  +ASAVAVPRYNK+HKVQNLLAY+V+KDGV+E+F+R L++TKAIKA ++D MM YM
Sbjct: 426 NKSQYIASAVAVPRYNKDHKVQNLLAYVVLKDGVEEQFERALDITKAIKADLQDVMMDYM 485

Query: 482 MPSKFLYRDSLPLTPNGKIDIKTLINEVNNR 512
            MPSKFLYR  LPLTPNGKIDIK L++EVN +
Sbjct: 486 MPSKFLYRKDLPLTPNGKIDIKGLMSEVNKK 516
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 374/510 (73%), Positives = 439/510 (85%)

Query: 4   MIHDMIKTIEHFAETQADFPVYDILGEVHTYGQLKVDSDSLAAHIDSLGLVEKSPVLVFG 63
            MI DMI +IE FA+TQADFPVYD LGE  TYGQLK DSDS+AA IDSL L+ KSPVLVFG
Sbjct: 1   MIKDMIDSIEQFAQTQADFPVYDCLGERRTYGQLKRDSDSIAAFIDSLALLAKSPVLVFG 60
```

```
Query: 64  GQEYEMLATFVALTKSGHAYIPVDQHSALDRIQAIMTVAQPSLIISIGEFPLEVDNVPIL 123
            Q Y+MLATFVALTKSGHAYIPVD HSA +RI AI+ +A+PSLII+I EFPL ++ + ++
Sbjct: 61  AQTYDMLATFVALTKSGHAYIPVDVHSAPERILAIIEIAKPSLIIAIEEFPLTIEGISLV 120

Query: 124 DVSQVSAIFEEKTPYEVTHSVKGDDNYYIIFTSGTTGLPKGVQISHDNLLSFTNWMISDD 183
            +S++ +    + PYE THSVKGDDNYYIIFTSGTTG PKGVQISHDNLLSFTNWMI D
Sbjct: 121 SLSEIESAKLAEMPYERTHSVKGDDNYYIIFTSGTTGQPKGVQISHDNLLSFTNWMIEDA 180

Query: 184 EFSVPERPQMLAQPPYSFDLSVMYWAPTLAMGGTLFALPKTVVNDFKKLFATINELPIQV 243
            F VP++PQMLAQPPYSFDLSVMYWAPTLA+GGTLFALPK +V DFK+LF TI +LP+ +
Sbjct: 181 AFDVPKQPQMLAQPPYSFDLSVMYWAPTLALGGTLFALPKELVADFKQLFTTIAQLPVGI 240

Query: 244 WTSTPSFADMALLSNDFNSETLPQLTHFYFDGEELTVKTAQKLRQRFPKARIVNAYGPTE 303
           WTSTPSFADMA+LS+DF    +P LTHFYFDGEELTV TA+KL +RFP A+I+NAYGPTE
Sbjct: 241 WTSTPSFADMAMLSDDFCQAKMPALTHFYFDGEELTVSTARKLFERFPSAKIINAYGPTE 300

Query: 304 ATVALSAVAITDEMLETCKRLPIGYTKDDSPTYVIDEEGHKLPNGEQGEIIIAGPAVSKG 363
           ATVALSA+ IT EM++   RLPIGY K DSPTY+IDE+G +L +GEQGEII+ GPAVSKG
Sbjct: 301 ATVALSAIEITREMVDNYTRLPIGYPKPDSPTYIIDEDGKELSSGEQGEIIVTGPAVSKG 360

Query: 364 YLNNPEKTAEAFFQFEGLPAYHTGDLGSMTDEGLLLYGGRMDFQIKFNGYRIELEDVSQN 423
           YLNNPEKTAEAFF F+G PAYHTGD+GS+T++ +LLYGGR+DFQIK+ GYRIELEDVSQ
Sbjct: 361 YLNNPEKTAEAFFTFKGQPAYHTGDIGSLTEDNILLYGGRLDFQIKYAGYRIELEDVSQQ 420

Query: 424 LNKSQYVKSAVAVPRYNKDHKVQNLLAYIVLKEGVRDDFERDLDLTKAIKEDLKDIMMDY 483
           LN+S  V SAVAVPRYNK+HKVQNLLAYIV+K+GV++ F+R+L+LTKAIK  +KD MM Y
Sbjct: 421 LNQSPMVASAVAVPRYNKEHKVQNLLAYIVVKDGVKERFDRELELTKAIKASVKDHMMSY 480

Query: 484 MMPSKFIYREDLPLTPNGKIDIKGLMSEVN 513
           MMPSKF+YR+ LPLTPNGKIDIK L++EVN
Sbjct: 481 MMPSKFLYRDSLPLTPNGKIDIKTLINEVN 510
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1851

A DNA sequence (GBSx1958) was identified in *S.agalactiae* <SEQ ID 5749> which encodes the amino acid sequence <SEQ ID 5750>. This protein is predicted to be a histidine protein kinase (phoR). Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-13.64   Transmembrane    9 -  25 (   5 -  32)
   INTEGRAL    Likelihood =-11.62   Transmembrane  136 - 152 ( 132 - 164)

----- Final Results -----
               bacterial membrane --- Certainty=0.6456(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB54569 GB:AJ006392 histidine kinase [Streptococcus pneumoniae]
 Identities = 105/416 (25%), Positives = 197/416 (47%), Gaps = 56/416 (13%)

Query: 7   KKFVFLTMSILIVVVLFLFAVSNRYNQYWDEYDAYRIVKLVAKNDY---LGIPGDEPIAL 63
           + F+F+ + +  ++V+ L  + NR +    +    ++  L+A  DY   L + G   I
Sbjct: 12  RDFIFILILLGFILVVTLLLLENRRDNIQLKQVNQKVKDLIA-GDYSKVLDMQGGSEITN 70

Query: 64  VTIDNQKMVKIQSNNTDLTNDVIEKSSLKL------LEQGKKSRKWKSFIYSIKE----- 112
           +T +   + ++      LT + +E+ S +L      +  G  +   +  I  I +
Sbjct: 71  ITNNLNDLSEV----IRLTQENLEQESKRLNSILFYMTDGVLATNRRGQIIMINDTAKKQ 126
```

```
Query: 113 ---YKDKTYTIAIMDLASYEVPYARRFLILVFT--------IFGFCLLAAVSLYLSR--- 158
               K+     +I++L   E  Y  R LI            I G  L   V   L R
Sbjct: 127 LGLVKEDVLNRSILELLKIEENYELRDLITQSPELLLDSQDINGEYLNLRVRFALIRRES 186

Query: 159 -FIVGPVE-----TEMTREKQ----FVSDASHELKTPIAAIRANVQVLEQ----QIPGNR 204
            FI G V      TE  +E++    FVS+ SHEL+TP+ ++++ ++ L++    +
Sbjct: 187 GFISGLVAVLHDTTEQEKEERERRLFVSNVSHELRTPLTSVKSYLEALDEGALCETVAPD 246

Query: 205 YLDHVVSETKRMEFLIEDLLNLSRLDEKRSKVNFKKLNLSVLCQEVLLTYESLAYEEEKC 264
           ++   + ET RM  ++ DLL+LSR+D   S ++ + +N +     +L  ++ +  +E++
Sbjct: 247 FIKVSLDETNRMMRMVTDLLHLSRIDNATSHLDVELINFTAFITFILNRFDKMKGQEKEK 306

Query: 265 LNDTIED----DVWIVGEESQIKQILIILLDNAIRHSLSKSAIQFSLKQARRKAILTISN 320
             + + D     +W+  +  ++ Q++  +L+NAI++S     I   +K    + IL+IS+
Sbjct: 307 KYELVRDYPINSIWMEIDTDKMTQVVDNILNNAIKYSPDGGKITVRMKTTEDQMILSISD 366

Query: 321 PSAIYSKEVMDNLFERFYQAKDDHADSLS---FGLGLSIAKAIVERHKGRIRAYQE 373
                 K+ +  +F+RFY+   D A S +    GLGLSIAK I+++HKG I A  E
Sbjct: 367 HGLGIPKQDLPRIFDRFYRV--DRARSRAQGGTGLGLSIAKEIIKQHKGFIWAKSE 420
```

A related sequence was also identified in GAS <SEQ ID 9131> which encodes the amino acid sequence <SEQ ID 9132>. Analysis of this protein sequence reveals the following:

```
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.30   Transmembrane    9 -  25 (   4 -  33)
   INTEGRAL    Likelihood =-10.35   Transmembrane  161 - 177 ( 154 - 190)
   PERIPHERAL  Likelihood =  4.35     142

----- Final Results -----
                bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 94/406 (23%), Positives = 190/406 (46%), Gaps = 31/406 (7%)

Query: 1   MFSDLRKKFVFLTMSILIVVVLFLFAVSNRYNQYWDEYDAYRIVKLVAKNDYLGIPGDEP 60
           MF+ +R +F+ +    + +++  +  + N    Y  + +  RI+ L++ N    +PG
Sbjct: 10  MFNRIRIRFIMIASIAIFIILSSIVGIINTARCYQSQQEINRILHLISSNKGK-LPGTTE 68

Query: 61  IAL-----VTIDNQKMVKIQS-----NNTDLTNDVIEKSSLKLLE------------QGK 98
            +      ++ D+    +  S     N   L+++    S+L   E            + K
Sbjct: 69  SSKRLGTKLSEDSLSQFRYYSVIFNANGHLLSSNTANISALDREEAQYFARLFAKSGEEK 128

Query: 99  KSRKWKSFIYS--IKEYKDKTYTIAIMDLASYEVPYARRFLILVFTIFG-FCLLAAVSLY 155
            S + +  +YS  I +  ++   + I+D   Y         + V   FG F     +
Sbjct: 129 GSYRHQDSVYSYLITQLPNEEKLVVILDTTFYFRSVGDLLAVSVMLAFGGFIFFVVLVSL 188

Query: 156 LSRFIVGPVETEMTREKQFVSDASHELKTPIAAIRANVQVLEQQIPGNRYLDHVVSETKR 215
            S  ++ P      ++++F+++A HELKTP+A I AN +++E     + +      + KR
Sbjct: 189 FSGMVIKPFVQNYEKQRRFITNAGHELKTPLAIISANNELVELMTGESEWTKSTSDQVKR 248

Query: 216 MEFLIEDLLNLSRLDEKRSKVNFKKLNLSVLCQEVLLTYESLAYEEEKCLNDTIEDDVWI 275
           +  LI  ++ L+RL+E+   V    ++ S + Q+    ++SL  ++ K  + TI+ ++ I
Sbjct: 249 LTGLINQMITLARLEEQPDVV-LHMVDFSAIAQDAAEDFKSLVLKDGKRFDLTIQPNIMI 307

Query: 276 VGEESQIKQILIILLDNAIRHSLSKSAIQFSLK---QARRKAILTISNPSAIYSKEVMDN 332
             EE  + +++ IL+DNA ++   K  ++ SL    + R++A L +SN
Sbjct: 308 KAEEKSLFELVTILVDNANKYCDPKGLVKVSLTTIGRRRKRAKLEVSNTYLEGKSIDYSR 367

Query: 333 LFERFYQAKDDH-ADSLSFGLGLSIAKAIVERHKGRIRAYQEKDQL 377
            FERFY+  + H +    +G+GLS+A+++V+  KG I    + D +
Sbjct: 368 FFERFYREDESHNSKEKGYGIGLSMAESMVKLFKGTITVNYKNDAI 413
```

A related GBS gene <SEQ ID 8915> and protein <SEQ ID 8916> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 7
McG: Discrim Score:     17.50
GvH: Signal Score (-7.5): -2.9
     Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 2 value: -13.64 threshold:  0.0
   INTEGRAL    Likelihood =-13.64   Transmembrane    9 -  25 (   5 -  32)
   INTEGRAL    Likelihood =-11.62   Transmembrane  136 - 152 ( 132 - 164)
   PERIPHERAL  Likelihood =  2.49     345
 modified ALOM score:    3.23

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6456(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
28.3/57.2% over 371aa
                                                 Listeria monocytogenes
  GP|6117973| LisK Insert characterized

ORF00341(631 - 1452 of 1785)
GP|6117973|gb|AAF03933.1|AF139908_3|AF139908(105   -   476   of   483)   LisK   {Listeria
monocytogenes}
%Match = 8.4
%Identity = 28.2  %Similarity = 57.1
Matches = 79  Mismatches = 113  Conservative Sub.s = 81

459       489       519       549       579       609       639       669
VKLVAKNDYLGIPGDEPIALVTIDNQKMVKIQSNNTDLTNDVIEKSSLKLLEQGKKSRKWKSFIYSIKEYKDKTYTIAIM
  :         |  : : | |        :  |         :|:  : |:    :   |   :  | |  ::  |  :
QGIGQMLLNEEEPEVKELLLATTSTLTNQDLTDNEEIKYLFNNDKTVNRKLQDQVINLYDKDGHFINKYYFSRSQDITSI
           50        60        70        80        90       100       110

699       729                                                 756
DLASYEVPYARRFLILVFTIFG-----------------------------------------FCLLAAVSLYLSRFI--
|:: | |    :|::   || |                                           || ||:|::|  :
DFSQYFVSGTDKFIMNKPTIDGQKMMTAQMPIVADDNTTVIGYAQVVNPLTSYNRMMDRLLVTMILLGAVALFISGMLGY
          130       140       150       160       170       180       190

783       813       843       873
-------------------------------------------------VGPVETEMTREKQFVSDASHELKTPIAAIRA
                                                 :  :||   ::|||| ||||||:||:  :
LLAQNFLNPLTRLARTMNDIRKNGFQKRIETKTNSRDEIGELTVVFNDMMTRIETSFEQQKQFVEDASHELRTPVQIMEG
          210       220       230       240       250       260       270

918       948       978      1008      1038      1068      1098
NVQVLEQ---QIPG--NRYLDHVVSETKRMEFLIEDLLNLSRLDEKRSKVNFKKLNLSVLCQEVLLTYESLAYEEEKCLN
::::| :     |   :  |:  ::| :||: |::::|:||| ::      ::  :::   ::|   :| : ||
HLKLLTRWGKDDPAVLDESLNASLTELERMKKLVQEMLDLSRAEQISQTKELQITDVNATVEQVRRNFE-VMYENFTFTL
          290       300       310       320       330       340       350

1128      1158      1188      1218      1248      1278      1308      1335
DTIEDDVWIVGEESQIKQILIILLDNAIRHSLSKSAIQFSLKQARRKAILTISNPSAIYSKEVMDNLFERFYQA-KDDHA
   : |:  : : ::::|||||::|||:::|   : :   : : :::  : : :     |:| :| :| |||:  |
KEDDTDLRALIQHNHLEQILIIIMDNAVKYSGDGTEVDMHVYKEQKQIHIDVRDYGEGISQEEIDKIFNRFYRVDKARSR
           370       380       390       400       410       420       430

1365      1395      1425      1452      1482      1512      1542      1572
DSLSFGLGLSIAKAIVERHKGRIRAYQEKDQ-LRLEVQLPIDGFWTNTMIN*RKNDETIFIFYW*NVIILRYFIVTNLLF
:    ||||:||| :|| : | | |  | |:   ::: ||     :
```

```
EKGGNGLGLAIAKQLVEGYLGTINAVSEPDKGTTIKITLPYIEPKSK
           450       460       470       480
```

SEQ ID 5750 (GBS34) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 15 (lane 9; MW 69kDa).

GBS34-GST was purified as shown in Figure 193, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1852

A DNA sequence (GBSx1959) was identified in *S.agalactiae* <SEQ ID 5753> which encodes the amino acid sequence <SEQ ID 5754>. This protein is predicted to be two-component response regulator (regX3). Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1986(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04091 GB:AP001508 two-component response regulator [Bacillus halodurans]
 Identities = 98/223 (43%), Positives = 145/223 (64%), Gaps = 5/223 (2%)

Query: 2   RLLVVEDEKSIAEAIQALLADKGYSVDLAFDGDDGLEYILTGLYDLVLLDIMLPKRSGLS 61
            R+L++EDEK IA  +Q  L  +GY  D AF G DGLE       +DLVLLD+MLP+ SGL
Sbjct: 3   RILIIEDEKKIARVLQLELEHEGYETDAAFSGSDGLETFQAHAWDLVLLDVMLPELSGLE 62

Query: 62  VLKRVREAGLETPIIFLTAKSQTYDKVNGLDLGADDYITKPFEADELLARIR--LRTRQS 119
            VL+R+R     TPII LTA++   DKV+GLDLGA+DYITKPFE +ELLAR+R  LRT Q+
Sbjct: 63  VLRRIRMTDPVTPIILLTARNSIPDKVSGLDLGANDYITKPFEIEELLARVRACLRTVQT 122

Query: 120 SLIRANQLRLGNIRLNTDSHELESKESSVKLSNKEFLLMEVFMRNAKQIIPKNQLISKVW 179
                 + L    + +N  + +++    +++L+ KEF L+  F++N  Q++ + Q+++ VW
Sbjct: 123 RERVEDTLMFQELTINEKTRDVQRGNETIELTPKEFELLVFFIKNKGQVLSREQILTNVW 182

Query: 180 GPSDNSEYNQLEVFISFLRKKLRFLKADIEIITTKGFGYSLEE 222
            G     + N ++V++ +LRKKL   +A   + T +G GY L+E
Sbjct: 183 GFDYYGDTNVIDVYVRYLRKKLSLTEA---LQTVRGVGYRLKE 222
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1853

A DNA sequence (GBSx1960) was identified in *S.agalactiae* <SEQ ID 5755> which encodes the amino acid sequence <SEQ ID 5756>. This protein is predicted to be 50S ribosomal protein L34-related protein. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5923(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC22660 GB:U32781 ribosomal protein L34 (rpL34) [Haemophilus influenzae Rd]
 Identities = 32/44 (72%), Positives = 37/44 (83%)

Query: 1   MKRTYQPSKIRRQRKHGFRHRMSTKNGRRVLASRRRKGRKVLSA 44
           MKRT+QPS ++R R HGFR RM+TKNGR+VLA RR KGRK LSA
Sbjct: 1   MKRTFQPSVLKRSRTHGFRARMATKNGRQVLARRRAKGRKSLSA 44
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5757> which encodes the amino acid sequence <SEQ ID 5758>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5385(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 42/44 (95%), Positives = 44/44 (99%)

Query: 1   MKRTYQPSKIRRQRKHGFRHRMSTKNGRRVLASRRRKGRKVLSA 44
           +KRTYQPSKIRRQRKHGFRHRMSTKNGRRVLA+RRRKGRKVLSA
Sbjct: 1   VKRTYQPSKIRRQRKHGFRHRMSTKNGRRVLAARRRKGRKVLSA 44
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1854

A DNA sequence (GBSx1961) was identified in *S.agalactiae* <SEQ ID 5759> which encodes the amino acid sequence <SEQ ID 5760>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -5.79   Transmembrane  122 - 138 ( 115 - 141)
   INTEGRAL    Likelihood = -4.35   Transmembrane   19 -  35 (  15 -  40)

----- Final Results -----
               bacterial membrane --- Certainty=0.3314(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF95990 GB:AE004350 conserved hypothetical protein [Vibrio cholerae]
 Identities = 79/145 (54%), Positives = 117/145 (80%)

Query: 1   MKTFVNNASKTVLSLWFGVMPTIMTVGTIALIISVSTPIFKILGTPFLPFLELLGIPEAD 60
           +++ +    +  + + FGV+P +M +GTIAL+I+  T +F +LG PF+PFLELLG+PEA
Sbjct: 314 VQSVIGEGIRNAVDMVFGVLPVVMGLGTIALVIAEYTSVFSLLGQPFIPFLELLGVPEAT 373

Query: 61  IASQTMIVGFSDMVVPSIMAAEIHSEMTRFIVATVSIVQLIYMSETGAVILGSKIPINIL 120
            AS+T++VGF+DM +P+I+AA I +EMTRF++A +S+ QLIYMSE GA++LGS+IP+NI+
Sbjct: 374 AASKTIVVGFADMFIPAILAASIDNEMTRFVIAAMSVTQLIYMSEVGALLLGSRIPVNIV 433

Query: 121 ELFIIFIERTIISLPIIVLMAHLFF 145
           ELF+IFI RT+I+LP+I  +AHL F
Sbjct: 434 ELFVIFILRTLITLPVIAAVAHLLF 458
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1855

A DNA sequence (GBSx1962) was identified in *S.agalactiae* <SEQ ID 5761> which encodes the amino acid sequence <SEQ ID 5762>. This protein is predicted to be D,D-carboxypeptidase (dacA-2). Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2443(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9485> which encodes amino acid sequence <SEQ ID 9486> was also identified. A further related GBS nucleic acid sequence <SEQ ID 10945> which encodes amino acid sequence <SEQ ID 10946> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA67776 GB:X99400 D,D-carboxypeptidase [Streptococcus pneumoniae]
 Identities = 193/383 (50%), Positives = 282/383 (73%), Gaps = 6/383 (1%)

Query: 1   MAVDLDSGKILYEKDANKPAAIASLTKIMTVYMVYKEIDNGNLKWNTKVNISDYPYQLTR 60
           +AV+ ++GKILYEKDA +P  IAS+TK++TVY+VY+ ++NG++  +T V+ISDYPYQLT
Sbjct: 33  IAVEANTGKILYEKDATQPVEIASITKLITVYLVYEALENGSITLSTPVDISDYPYQLTT 92

Query: 61  ESDASNVPLEKRRYTVKQLVDAAMISSANSAAIALAEHISGTESKFVDKMTAQLEKWGIH 120
            S+ASN+P+E R YTV++L++A ++SSANSAAIALAE I+G+E  FVD M A+L +WGI
Sbjct: 93  NSEASNIPMEARNYTVEELLEATLVSSANSAAIALAEKIAGSEKDFVDMMRAKLLEWGIQ 152

Query: 121 DSHLVNASGLNNSMLGNHIYPKSSQNDENKMSARDIAIVAYHLVNEYPSILKITSKSVAK 180
           D+ +VN +GLNN  LG++IYP S +++ENK+SA D+AIVA +L+ +YP +L+IT K  +
Sbjct: 153 DATVVNTTGLNNETLGDNIYPGSKKDEENKLSAYDVAIVARNLIKKYPQVLEITKKPSST 212

Query: 181 FDKDIMHSYNYMLPDMPVFRPGITGLKTGTTELAGQSFIATSTESGMRLLTVIMHADKAD 240
           F    + S NYML  MP +R G  GLKTGTT+ AG+SF+ T+ E GMR++TV+++AD  D
Sbjct: 213 FAGMTITSTNYMLEGMPAYRGGFDGLKTGTTDKAGESFVGTTVEKGMRVITVVLNADHQD 272

Query: 241 KDKYARFTATNSLLNYITNTYEPNLVLAKGAAYKGKEASVRDGKEQSVIAVAKNDLKVVQ 300
            + YARFTAT+SL++YI++T+    ++ +G AY+  +A V+DGKE +VIAVA  D+ +++
Sbjct: 273 NNPYARFTATSSLMDYISSTFTLRKIVQQGDAYQDSKAPVQDGKEDTVIAVAPEDIYLIE 332

Query: 301 KKNITKQNQLKINF---KKELTAPITKKENLGKAYYVDLNKVGKGYLIKE-PSVHLVAKD 356
           +  +  Q+   + F    K + AP+     +G   Y D + +G+GY+  E PS  +VA
Sbjct: 333 R--VGNQSSQSVQFTPDSKAIPAPLEAGTVVGHLTYEDKDLIGQGYITTERPSFEMVADK 390

Query: 357 SIERSFFLKVWWNHFVRYVNEKL 379
            IE++FFLKVWWN FVR+VNEKL
Sbjct: 391 KIEKAFFLKVWWNQFVRFVNEKL 413
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5763> which encodes the amino acid sequence <SEQ ID 5764>. Analysis of this protein sequence reveals the following:

```
     Possible site: 21

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
```

```
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 176/380 (46%), Positives = 257/380 (67%), Gaps = 3/380 (0%)

Query: 1   MAVDLDSGKILYEKDANKPAAIASLTKIMTVYMVYKEIDNGNLKWNTKVNISDYPYQLTR 60
           +AVDL+SGK+LYEKDA +   +AS++K++T Y+VYKE+  G L W++ V IS+YPY+LT
Sbjct: 33  IAVDLESGKVLYEKDAKEVVPVASVSKLLTTYLVYKEVSKGKLNWDSPVTISNYPYELTT 92

Query: 61  ESDASNVPLEKRRYTVKQLVDAAMISSANSAAIALAEHISGTESKFVDKMTAQLEKWGIH 120
               SNVPL+KR+YTVK+L+ A ++++ANS AIALAE I GTE KFVDKM  QL +WGI
Sbjct: 93  NYTISNVPLDKRKYTVKELLSALVVNNANSPAIALAEKIGGTEPKFVDKMKKQLRQWGIS 152

Query: 121 DSHLVNASGLNNSMLGNHIYPKSSQNDENKMSARDIAIVAYHLVNEYPSILKITSKSVAK 180
           D+ +VN++GL N  LG + YP +  +DEN   A D+AI+A HL+ E+P +LK++SKS
Sbjct: 153 DAKVVNSTGLTNHFLGANTYPNTEPDDENCFCATDLAIIARHLLLEFPEVLKLSSKSSTI 212

Query: 181 FDKDIMHSYNYMLPDMPVFRPGITGLKTGTTELAGQSFIATSTESGMRLLTVIMHADKAD 240
           F    ++SYNYML  MP +R G+ GL  G ++ AG SF+ATS E+ MR++TV+++AD++
Sbjct: 213 FAGQTIYSYNYMLKGMPCYREGVDGLFVGYSKKAGASFVATSVENQMRVITVVLNADQSH 272

Query: 241 KDKYARFTATNSLLNYITNTYEPNLVLAKGAAYKGKEASVRDGKEQSVIAVAKNDLKVVQ 300
           +D  A F  TN LL Y+   ++   ++        K   V D  E++V  VA+N L  ++
Sbjct: 273 EDDLAIFKTTNQLLQYLLINFQKVQLIENNKPV--KTLYVLDSPEKTVKLVAQNSLFFIK 330

Query: 301 KKNITKQNQLKINFKKE-LTAPITKKENLGKAYYVDLNKVGKGYLIKEPSVHLVAKDSIE 359
             +   +N + I  K   + AP++K + LG+A   D + +G+GYL   PS++L+ + +I
Sbjct: 331 PIHTKTKNTVHITKKSSTMIAPLSKGQVLGRATLQDKHLIGQGYLDTPPSINLILQKNIS 390

Query: 360 RSFFLKVWWNHFVRYVNEKL 379
           +SFFLKVWWN FVRYVN  L
Sbjct: 391 KSFFLKVWWNRFVRYVNTSL 410
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1856

A DNA sequence (GBSx1963) was identified in *S.agalactiae* <SEQ ID 5765> which encodes the amino acid sequence <SEQ ID 5766>. This protein is predicted to be penicillin binding protein 4 (pdp4) (dacA-1). Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-12.58    Transmembrane  368 - 384 ( 363 - 394)

----- Final Results -----
                bacterial membrane --- Certainty=0.6031(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA60582 GB:X87104 penicillin binding protein 4 [Staphylococcus
           aureus]
 Identities = 117/333 (35%), Positives = 188/333 (56%), Gaps = 8/333 (2%)

Query: 5   IVSFLCILLSLTCVNSVQAEEHKDIMQITREAGY-DVKDINKPKASIVIDNKGHILWEDN 63
           I+  LC+ LS+    +  A      +Q   + GY  +    +P +++ +   G +L++ N
Sbjct: 7   IIIILCLTLSIMTPYAQAANSDVTPVQAANQYGYAGLSAAYEPTSAVNVSQTGQLLYQYN 66

Query: 64  ADLERDPASMSKMFTLYLLFEDLAKGKTSLNTTVTATETDQAISKIYEISNNNIHAGVAY 123
            D + +PASM+K+ T+YL  E + KG+ SL+ TVT T  +  +S + E+SN  ++ G  +
```

```
Sbjct: 67  IDTKWNPASMTKLMTMYLTLEAVNKGQLSLDDTVTMTNKEYIMSTLPELSNTKLYPGQVW 126

Query: 124 PIRELITMTAVPSSNVATIMIANHLSQNNPDAFIKRINETAKKLGMTKTHFYNPSGAVAS 183
            I +L+ +T   SSN A +++A  +S+N  D F+  +N  AK +GM  THF NP+GA  S
Sbjct: 127 TIADLLQITVSNSSNAAALILAKKVSKNTSD-FVDLMNNKAKAIGMKNTHFVNPTGAENS 185

Query: 184 AFNGLYSPKEYDNNATNVTTARDLSILTYHFLKKYPDILNYTKYPEVKAMVGTPYEETFT 243
                ++P +Y +    VTTARD +IL  H +K+ P IL++T     K +  T +  T+
Sbjct: 186 RLR-TFAPTKYKDQERTVTTARDYAILDLHVIKETPKILDFT-----KQLAPTTHAVTYY 239

Query: 244 TYNYSTPGAKFGLEGVDGLKTGSSPSAAFNALVTAKRQNTRLITVVLGVGDWSDQDGEYY 303
           T+N+S  GAK  L G DGLKTGSS +A +N  +T KR   R+  V++G GD+ +  GE
Sbjct: 240 TFNFSLEGAKMSLPGTDGLKTGSSDTANYNHTITTKRGKFRINQVIMGAGDYKNLGGEKQ 299

Query: 304 RHPFVNALVEKGFKDAKNISSKTPVLKAVKPKK 336
           R+   NAL+E+ F   K +   +   + +  KK
Sbjct: 300 RNMMGNALMERSFDQYKYVKILSKGEQRINGKK 332
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5767> which encodes the amino acid sequence <SEQ ID 5768>. Analysis of this protein sequence reveals the following:

```
     Possible site: 23

  >>> Seems to have a cleavable N-term signal seq.
     INTEGRAL    Likelihood =-15.18    Transmembrane  371 - 387 ( 364 - 392)

 ----- Final Results -----
                  bacterial membrane --- Certainty=0.7071(Affirmative) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA62899 GB:X91786 penicillin-binding protein 4 [Staphylococcus
           aureus]
 Identities = 119/328 (36%), Positives = 184/328 (55%), Gaps = 19/328 (5%)

Query: 6   ILTIFTFICF--SVMPLVHAEDVMDIT-----RQAGYT-VSEVNRPKSSIVVDANSSDIL 57
           +++I   +C   S+M         D+T      Q GY  +S    P S++ V + +  +L
Sbjct: 4   LISIIIILCLTLSIMTPYAQATNSDVTPVQAANQYGYAGLSAAYEPTSAVNV-SQTGQLL 62

Query: 58  WQDNIDIPRDPASMSKMFTLYILFEELAKGKITMDTTITATPTDQAIANIYEISNNNIVA 117
           +Q NID   +PASM+K+ T+Y+  E + KG++++D T+T T  +  ++ + E+SN  +
Sbjct: 63  YQYNIDTKWNPASMTKLMTMYLTLEAVNKGQLSLDDTVTMTNKEYIMSTLPELSNTKLYP 122

Query: 118 GVAYPIRDLITMTAVPSSNAATVMIANYLSNNDASAFIDRVNATAKQLGMTNTHFSNASG 177
           G  + I DL+ +T   SSNAA +++A  +S N  S F+D +N  AK +GM NTHF N +G
Sbjct: 123 GQVWTIADLLQITVSNSSNAAALILAKKVSKN-TSDFVDLMNNKAKAIGMKNTHFVNPTG 181

Query: 178 AAAQAFQGYYNPTKYDLSASNITTARDLSKLLYAFLKKYPEIISFTNKSVVHTMVGTPYE 237
           A     +  + PTKY      +TTARD + L    +K+ P+I+ FT +    T+  T
Sbjct: 182 AENSRLR-TFAPTKYKDQERTVTTARDYAILDLHVIKETPKILDFTKQLAPTTLAVT--- 237

Query: 238 EEFHTYNHSLPDNQFGMKGVDGLKTGSSPSAAFNAMITAKRGKTRLITIVMGVGDWSDQN 297
             ++T+N SL   +  + G DGLKTGSS +A +N  IT KRGK R+  ++MG GD+ +
Sbjct: 238 --YYTFNFSLEGAKMSLPGTDGLKTGSSDTANYNHTITTKRGKFRINQVIMGAGDYKNLG 295

Query: 298 GEFYRHPFVNALTEKGF---KDSKTLSK 322
           GE  R+   NAL E+ F   K  K LSK
Sbjct: 296 GEKQRNMMGNALMERSFDQYKYVKILSK 323
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 226/382 (59%), Positives = 289/382 (75%), Gaps = 7/382 (1%)

Query: 12  LLSLTCVNSVQAEEHKDIMQITREAGYDVKDINKPKASIVID-NKGHILWEDNADLERDP 70
           + +  C + +     +D+M ITR+AGY V ++N+PK+SIV+D N   ILW+DN D+ RDP
```

```
Sbjct: 9   IFTFICFSVMPLVHAEDVMDITRQAGYTVSEVNRPKSSIVVDANSSDILWQDNIDIPRDP 68

Query: 71  ASMSKMFTLYLLFEDLAKGKTSLNTTVTATETDQAISKIYEISNNNIHAGVAYPIRELIT 130
           ASMSKMFTLY+LFE+LAKGK +++TT+TAT TDQAI+ IYEISNNNI AGVAYPIR+LIT
Sbjct: 69  ASMSKMFTLYILFEELAKGKITMDTTITATPTDQAIANIYEISNNNIVAGVAYPIRDLIT 128

Query: 131 MTAVPSSNVATIMIANHLSQNNPDAFIKRINETAKKLGMTKTHFYNPSGAVASAFNGLYS 190
           MTAVPSSN AT+MIAN+LS N+  AFI R+N TAK+LGMT THF N SGA A AF G Y+
Sbjct: 129 MTAVPSSNAATVMIANYLSNNDASAFIDRVNATAKQLGMTNTHFSNASGAAAQAFQGYYN 188

Query: 191 PKEYDNNATNVTTARDLSILTYHFLKKYPDILNYTKYPEVKAMVGTPYEETFTTYNYSTP 250
           P +YD +A+N+TTARDLS L Y FLKKYP+I+++T    V  MVGTPYEE F TYN+S P
Sbjct: 189 PTKYDLSASNITTARDLSKLLYAFLKKYPEIISFTNKSVVHTMVGTPYEEEFHTYNHSLP 248

Query: 251 GAKFGLEGVDGLKTGSSPSAAFNALVTAKRQNTRLITVVLGVGDWSDQDGEYYRHPFVNA 310
             +FG++GVDGLKTGSSPSAAFNA++TAKR  TRLIT+V+GVGDWSDQ+GE+YRHPFVNA
Sbjct: 249 DNQFGMKGVDGLKTGSSPSAAFNAMITAKRGKTRLITIVMGVGDWSDQNGEFYRHPFVNA 308

Query: 311 LVEKGFKDAKNISSKT-PVLKAVKPKKEVTKTKTKSIQE--QPQTKEQWWTKTDQFIQSH 367
           L EKGFKD+K +S K    L+ + P+   TK +T S Q+  +   K+ +  + + F+  +
Sbjct: 309 LTEKGFKDSKTLSKKARQKLEKLVPQ---TKKETSSKQQHFKATKKQSYLERVEDFMNHN 365

Query: 368 FVSILIVLGTIAILCLLAGIVL 389
              +LI L    I  LL  +V+
Sbjct: 366 HTFLLICLAIFIITILLLSLVV 387
```

A related GBS gene <SEQ ID 8917> and protein <SEQ ID 8918> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 9
McG: Discrim Score:    -14.02
GvH: Signal Score (-7.5): -2.54
     Possible site: 60
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value: -12.58 threshold:  0.0
   INTEGRAL    Likelihood =-12.58   Transmembrane  339 - 355 ( 334 - 365)
   PERIPHERAL  Likelihood =  1.38      99
 modified ALOM score:   3.02

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6031(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01254(301 - 1386 of 1698)
EGAD|40430|42591(32 - 419 of 431) penicillin binding protein 4 (pdp4) {Staphylococcus
aureus} GP|1125682|emb|CAA60585.1||X87105 penicillin binding protein 4 {Staphylococcus
aureus} GP|1125686|emb|CAA60582.1||X87104 penicillin binding protein 4 {Staphylococcus
aureus}
%Match = 17.3
%Identity = 36.3  %Similarity = 59.6
Matches = 123  Mismatches = 130  Conservative Sub.s = 79

264       294       324       351       381       411       441       471
FPLHFIIPDLCKLCAS*RHKDIMQITREAGY-DVKDINKPKASIVIDNKGHILWEDNADLERDPASMSKMFTLYLLFEDL
                      :|   : ||  :    :| ::: :   |::|:: | | : ||||:|: |:|| :| :
ILCLTLSIMTPYAQAANSDVTPVQAANQYGYAGLSAAYEPTSAVNVSQTGQLLYQYNIDTKWNPASMTKLMTMYLTLEAV
         20        30        40        50        60        70        80

501       531       561       591       621       651       681       711
AKGKTSLNTTVTATETDQAISKIYEISNNNIHAGVAYPIRELITMTAVPSSNVATIMIANHLSQNNPDAFIKRINETAKK
 ||: ||: ||| |  :  :| : |:||  :: |  : | :|: :|   ||| | :::|  :|:|  | |:  :|  ||
NKGQLSLDDTVTMTNKEYIMSTLPELSNTKLYPGQVWTIADLLQITVSNSSNAAALILAKKVSKNTSD-FVDLMNNKAKA
```

```
        100       110       120       130       140       150        160

741       771       801       831       861       891       921       951
LGMTKTHFYNPSGAVASAFNGLYSPKEYDNNATNVTTARDLSILTYHFLKKYPDILNYTKYPEVKAMVGTPYEETFTTYN
:||  ||| ||:||  |     ::| :| :    |||||| :||  | :|: | ||::||      :  | :  |: |:|
IGMKNTHFVNPTGAENSR-LRTFAPTKYKDQERTVTTARDYAILDLHVIKETPKILDFTK-----QLAPTTHAVTYYTFN
         180        190       200       210       220            230       240

981      1011      1041      1071      1101      1131      1161
YSTPGAKFGLEGVDGLKTGSSPSAAFNALVTAKRQNTRLITVVLGVGDWSDQDGEYYRHPFVNALVEKGFKDAK------
:|  |||  | | |||||||| :| :|  :| ||   |:  |::| ||: :  ||  |:   |||:|: |   |
FSLEGAKMSLPGTDGLKTGSSDTANYNHTITTKRGKFRINQVIMGAGDYKNLGGEKQRNMMGNALMERSFDQYKYVKILS
               260       270       280       290       300       310       320

1179      1209      1239      1266
------------------------------------------------NISSKTPVLKAVKPKKEVTKTKTKSI-QEQPQ
                                                |   : | ::  :|  :   |  ||: :|:|
KGEQRINGKKYYVENDLYDVLPSDFSKKDYKLVVEDGKVHADYPREFINKDYRPPTVEVHQPIIQKANTVAKSMWEEHP-
               340       350       360       370       380       390       400

1296      1326      1356      1386      1416      1446      1476      1506
TKEQWWTKTDQFIQSHFVSILIVLGTIAILCLLAGIVLLIKRSR**LC*YKSPLHQ*HRGFLLSLEIFN*PTEPSIS*EI
                      ::  |   ||:||: |::
----------------------LFTIIGGACLVAGLALIVHMIINRLFRKRK
                            410       420       430
```

SEQ ID 8918 (GBS379) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 68 (lane 5; MW 44kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 3; MW 68.9kDa).

GBS379-GST was purified as shown in Figure 212, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1857

A DNA sequence (GBSx1964) was identified in *S.agalactiae* <SEQ ID 5769> which encodes the amino acid sequence <SEQ ID 5770>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4039(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15256 GB:Z99120 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 316/459 (68%), Positives = 386/459 (83%)

Query: 14  DLGEYKFGFHDDVKPIYSTGKGLNEAVIRELSAAKGEPEWMLDFRLKSLETFNKMPMQTW 73
           D+GEYK+GFHD    I+ + +GL + ++ E+S  K EP+WMLDFRLKSLE F  MPM  W
Sbjct: 7   DIGEYKYGFHDKDVSIFRSERGLTKEIVEEISRMKEEPQWMLDFRLKSLEHFYNMPMPQW 66

Query: 74  GADLSDIDFDDIIYYQKASDKPARDWDDVPEKIKETFERIGIPEAERAYLAGASAQYESE 133
           G DL+ ++FD+I YY K S++  R WD+VPE+IK+TF+++GIPEAE+ YLAG SAQYESE
Sbjct: 67  GGDLNSLNFDEITYYVKPSERSERSWDEVPEEIKQTFDKLGIPEAEQKYLAGVSAQYESE 126

Query: 134 VVYHNMKEEYDKLGIVFTDTDSALKEYPELFKKYFAKLVPPTDNKLAALNSAVWSGGTFI 193
           VVYHNMKE+ +  GIVF DTDSALKE  ++F++++AK++PPTDNK AALNSAVWSGG+FI
Sbjct: 127 VVYHNMKEDLEAQGIVFKDTDSALKENEDIFREHWAKVIPPTDNKFAALNSAVWSGGSFI 186
```

```
Query: 194 YVPKGVKVDIPLQTYFRINNENTGQFERTLIIVDEGASVHYVEGCTAPTYSSNSLHAAIV 253
           YVPKGVKV+ PLQ YFRIN+EN GQFERTLIIVDE ASVHYVEGCTAP Y++NSLH+A+V
Sbjct: 187 YVPKGVKVETPLQAYFRINSENMGQFERTLIIVDEEASVHYVEGCTAPVYTTNSLHSAVV 246

Query: 254 EIFALDGAYMRYTTIQNWSDNVYNLVTKRATAKKDATVEWIDGNLGAKTTMKYPSVYLDG 313
           EI    G Y RYTTIQNW++NVYNLVTKR   +++AT+EWIDGN+G+K TMKYP+  L G
Sbjct: 247 EIIVKKGGYCRYTTIQNWANNVYNLVTKRTVCEENATMEWIDGNIGSKLTMKYPACILKG 306

Query: 314 EGARGTMLSIAFANKGQHQDTGAKMIHNAPHTSSSIVSKSIAKGGGKVDYRGQVTFNKDS 373
           EGARG  LSIA A KGQHQD GAKMIH AP+TSS+IVSKSI+K GGKV YRG V F + +
Sbjct: 307 EGARGMTLSIALAGKGQHQDAGAKMIHLAPNTSSTIVSKSISKQGGKVTYRGIVHFGRKA 366

Query: 374 KKSVSHIECDTILMDDISKSDTIPFNEIHNSQVALEHEAKVSKISEEQLYYLMSRGLSEA 433
           + + S+IECDT++MD+ S SDTIP+NEI N  ++LEHEAKVSK+SEEQL+YLMSRG+SE
Sbjct: 367 EGARSNIECDTLIMDNKSTSDTIPYNEILNDNISLEHEAKVSKVSEEQLFYLMSRGISEE 426

Query: 434 EATEMIVMGFVEPFTKELPMEYAVELNRLISYEMEGSVG 472
           EATEMIVMGF+EPFTKELPMEYAVE+NRLI +EMEGS+G
Sbjct: 427 EATEMIVMGFIEPFTKELPMEYAVEMNRLIKFEMEGSIG 465
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5771> which encodes the amino acid sequence <SEQ ID 5772>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3780(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 445/472 (94%), Positives = 461/472 (97%)

Query: 1   MSEINEKVEPQPIDLGEYKFGFHDDVKPIYSTGKGLNEAVIRELSAAKGEPEWMLDFRLK 60
           MS+INEKVEP+PIDLG+Y+FGFHDDV+PIYSTGKGL+EAV+RELSAAK EPEWML+FRLK
Sbjct: 1   MSDINEKVEPKPIDLGDYQFGFHDDVEPIYSTGKGLSEAVVRELSAAKNEPEWMLEFRLK 60

Query: 61  SLETFNKMPMQTWGADLSDIDFDDIIYYQKASDKPARDWDDVPEKIKETFERIGIPEAER 120
           SLETFNKMPMQTWGADLSDI+FDDIIYYQKASDKPAR WDDVPEKIKETF+RIGIPEAER
Sbjct: 61  SLETFNKMPMQTWGADLSDINFDDIIYYQKASDKPARSWDDVPEKIKETFDRIGIPEAER 120

Query: 121 AYLAGASAQYESEVVYHNMKEEYDKLGIVFTDTDSALKEYPELFKKYFAKLVPPTDNKLA 180
           AYLAGASAQYESEVVYHNMK E++KLGI+FTDTDSALKEYP+LFK+YFAKLVPPTDNKLA
Sbjct: 121 AYLAGASAQYESEVVYHNMKGEFEKLGIIFTDTDSALKEYPDLFKQYFAKLVPPTDNKLA 180

Query: 181 ALNSAVWSGGTFIYVPKGVKVDIPLQTYFRINNENTGQFERTLIIVDEGASVHYVEGCTA 240
           ALNSA WSGGTFIYVPKGVKVDIPLQTYFRINNENTGQFERTLIIVDEGASVHYVEGCTA
Sbjct: 181 ALNSAAWSGGTFIYVPKGVKVDIPLQTYFRINNENTGQFERTLIIVDEGASVHYVEGCTA 240

Query: 241 PTYSSNSLHAAIVEIFALDGAYMRYTTIQNWSDNVYNLVTKRATAKKDATVEWIDGNLGA 300
           PTYSSNSLHAAIVEIFALDGAYMRYTTIQNWSDNVYNLVTKRA A  DATVEWIDGNLGA
Sbjct: 241 PTYSSNSLHAAIVEIFALDGAYMRYTTIQNWSDNVYNLVTKRARALTDATVEWIDGNLGA 300

Query: 301 KTTMKYPSVYLDGEGARGTMLSIAFANKGQHQDTGAKMIHNAPHTSSSIVSKSIAKGGGK 360
           KTTMKYPSVYLDG GARGTMLSIAFAN GQHQDTGAKMIHNAPHTSSSIVSKSIAK GGK
Sbjct: 301 KTTMKYPSVYLDGPGARGTMLSIAFANAGQHQDTGAKMIHNAPHTSSSIVSKSIAKSGGK 360

Query: 361 VDYRGQVTFNKDSKKSVSHIECDTILMDDISKSDTIPFNEIHNSQVALEHEAKVSKISEE 420
           VDYRGQVTFNK SKKSVSHIECDTILMDDISKSDTIPFNEIHNSQVALEHEAKVSKISEE
Sbjct: 361 VDYRGQVTFNKQSKKSVSHIECDTILMDDISKSDTIPFNEIHNSQVALEHEAKVSKISEE 420

Query: 421 QLYYLMSRGLSEAEATEMIVMGFVEPFTKELPMEYAVELNRLISYEMEGSVG 472
           QLYYLMSRGLSE+EATEMIVMGFVEPFTKELPMEYAVELNRLISYEMEGSVG
Sbjct: 421 QLYYLMSRGLSESEATEMIVMGFVEPFTKELPMEYAVELNRLISYEMEGSVG 472
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1858

A DNA sequence (GBSx1965) was identified in *S.agalactiae* <SEQ ID 5773> which encodes the amino acid sequence <SEQ ID 5774>. This protein is predicted to be nitrogen fixation protein (nifU). Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1078(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15257 GB:Z99120 similar to NifU protein homolog [Bacillus subtilis]
 Identities = 72/139 (51%), Positives = 92/139 (65%)

Query: 4   SKLDNLYMAVVADHSKHPHHHGFLEGVEQVQLNNPTCGDVISLSVKFDGNIISDIAFAGN 63
           + LD LY  V+ DH K+P + G L     V +NNPTCGD I L++K DG+I+ D  F G
Sbjct: 5   ANLDTLYRQVIMDHYKNPRNKGVLNDSIVVDMNNPTCGDRIRLTMKLDGDIVEDAKFEGE 64

Query: 64  GCTISTASSSMMTDAVIGKTKEEALQLADVFSKMVQGDQNPKQEKLGDAEFLAGVSKFPQ 123
           GC+IS AS+SMMT A+ GK  E AL ++ +FS M+QG +      LGD E L GVSKFP
Sbjct: 65  GCSISMASASMMTQAIKGKDIETALSMSKIFSDMMQGKEYDDSIDLGDIEALQGVSKFPA 124

Query: 124 RIKCATLSWNALRKAIERD 142
           RIKCATLSW AL K + ++
Sbjct: 125 RIKCATLSWKALEKGVAKE 143
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5775> which encodes the amino acid sequence <SEQ ID 5776>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1202(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 114/146 (78%), Positives = 133/146 (91%)

Query: 1   MALSKLDNLYMAVVADHSKHPHHHGFLEGVEQVQLNNPTCGDVISLSVKFDGNIISDIAF 60
           MALSKL++LYMAVVADHSK PHHHG L+GVE VQLNNPTCGDVISL+VKFD + I DIAF
Sbjct: 1   MALSKLNHLYMAVVADHSKRPHHHGQLDGVEAVQLNNPTCGDVISLTVKFDEDKIEDIAF 60

Query: 61  AGNGCTISTASSSMMTDAVIGKTKEEALQLADVFSKMVQGDQNPKQEKLGDAEFLAGVSK 120
           AGNGCTISTASSSMMTDAVIGK+KEEAL LAD+FS+MVQG +NP Q++LG+AE LAGV+K
Sbjct: 61  AGNGCTISTASSSMMTDAVIGKSKEEALALADIFSEMVQGQENPAQKELGEAELLAGVAK 120

Query: 121 FPQRIKCATLSWNALRKAIERDNQAE 146
           FPQRIKC+TL+WNAL++AI+R   A+
Sbjct: 121 FPQRIKCSTLAWNALKEAIKRSANAQ 146
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1859

A DNA sequence (GBSx1966) was identified in *S.agalactiae* <SEQ ID 5777> which encodes the amino acid sequence <SEQ ID 5778>. This protein is predicted to be nitrogen fixation protein (nifS) (b1680). Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2453(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15258 GB:Z99120 similar to NifS protein homolog [Bacillus subtilis]
 Identities = 240/400 (60%), Positives = 306/400 (76%), Gaps = 5/400 (1%)

Query: 9    LKQDFPILNQLVNDEPLIYLDNAATTQKPNQVLEALRDYYQNDNANVHRGVHTLAERATA 68
            +++ FPIL+Q VN   L+YLD+AAT+QKP  V+E L  YY   N+NVHRGVHTL  RAT
Sbjct: 6    IREQFPILHQQVNGHDLVYLDSAATSQKPRAVIETLDKYYNQYNSNVHRGVHTLGTRATD 65

Query: 69   QYENAREKARQFLNAKLSKEILFTRGTTTGLNWVA-KFAESILERGDEVLISIMEHHSNI 127
             YE AREK R+F+NAK   EI+FT+GTTT LN VA  +A + L+ GDEV+I+ MEHH+NI
Sbjct: 66   GYEGAREKVRKFINAKSMAEIIFTKGTTTSLNMVALSYARANLKPGDEVVITYMEHHANI 125

Query: 128 IPWQQACERTGAKLVYAYLK-DGSLDLEDFYNKLSSKTKFVSLAHISNVLGCVTPVKAIA 186
            IPWQQA + TGA L Y  L+ DG++ LED    ++S TK V+++H+SNVLG V P+K +A
Sbjct: 126 IPWQQAVKATGATLKYIPLQEDGTISLEDVRETVTSNTKIVAVSHVSNVLGTVNPIKEMA 185

Query: 187 ERVHQVGAYMVVDGAQSAPHMAIDVQDLDCDFFALSGHKMLGPTGIGVLYGKESILDKMP 246
            +  H  GA +VVDGAQS PHM IDVQDLDCDFFALS HKM GPTG+GVLYGK+++L+ M
Sbjct: 186 KIAHDNGAVIVVDGAQSTPHMKIDVQDLDCDFFALSSHKMCGPTGVGVLYGKKALLENME 245

Query: 247 PVEFGGEMIDFVYEQSATWKELPWKFEAGTPNIAGAIAFGEALDYLTDVGMDEIHQYEQS 306
            P EFGGEMIDFV    +TWKELPWKFEAGTP IAGAI  G A+D+L ++G+DEI ++E
Sbjct: 246 PAEFGGEMIDFVGLYESTWKELPWKFEAGTPIIAGAIGLGAAIDFLEEIGLDEISRHEHK 305

Query: 307 LVSYVLPKLQAIDGLTIYGPSDAESHVGVIAFNLEGLHPHDVATAMDYEGVAVRAGHHCA 366
            L +Y L + + +DG+T+YGP   E   G++ FNL+ +HPHDVAT +D EG+AVRAGHHCA
Sbjct: 306 LAAYALERFRQLDGVTVYGP---EERAGLVTFNLDDVHPHDVATVLDAEGIAVRAGHHCA 362

Query: 367 QPLINHLGIHSAVRASFYFYNTKEDCDKLVDAIQKTKEFF 406
            QPL+  L + +  RASFY YNT+E+ DKLV+A+QKTKE+F
Sbjct: 363 QPLMKWLDVTATARASFYLYNTEEEIDKLVEALQKTKEYF 402
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5779> which encodes the amino acid sequence <SEQ ID 5780>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3714(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 293/408 (71%), Positives = 349/408 (84%)

Query: 3    LLDSYKLKQDFPILNQLVNDEPLIYLDNAATTQKPNQVLEALRDYYQNDNANVHRGVHTL 62
            LLD+  +KQDF ILNQ VNDEPL+YLDNAATTQKP  VLEAL+ YYQ DNANVHRGVHTL
Sbjct: 1    LLDAKDIKQDFQILNQQVNDEPLVYLDNAATTQKPALVLEALQSYYQEDNANVHRGVHTL 60
```

```
Query: 63  AERATAQYENAREKARQFLNAKLSKEILFTRGTTTGLNWVAKFAESILERGDEVLISIME 122
           AERAT +YE +R++   F++AK SKE+LFTRGTTT LNWVA+FAE +L   DEVLISIME
Sbjct: 61  AERATLKYEASRQQVADFIHAKSSKEVLFTRGTTTSLNWVARFAEQVLTPEDEVLISIME 120

Query: 123 HHSNIIPWQQACERTGAKLVYAYLKDGSLDLEDFYNKLSSKTKFVSLAHISNVLGCVTPV 182
           HH+NIIPWQQAC++TGA+LVY YLKDG LD++D  NKL++KT+FVSL H+SNVLGC+ P+
Sbjct: 121 HHANIIPWQQACQKTGARLVYVYLKDGQLDMDDLANKLTTKTRFVSLVHVSNVLGCINPI 180

Query: 183 KAIAERVHQVGAYMVVDGAQSAPHMAIDVQDLDCDFFALSGHKMLGPTGIGVLYGKESIL 242
           K IA+  H  GAY+VVDGAQS PH+AIDVQDLDCDFFA S HKMLGPTG+GVLYGKE +L
Sbjct: 181 KEIAKLAHAKGAYLVVDGAQSVPHLAIDVQDLDCDFFAFSAHKMLGPTGLGVLYGKEELL 240

Query: 243 DKMPPVEFGGEMIDFVYEQSATWKELPWKFEAGTPNIAGAIAFGEALDYLTDVGMDEIHQ 302
           +++ P+EFGGEMIDFVYEQ ATWKELPWKFEAGTP+IAGAI    A+ YL  +GM +IH
Sbjct: 241 NQVEPLEFGGEMIDFVYEQEATWKELPWKFEAGTPHIAGAIGLSAAISYLQRLGMADIHA 300

Query: 303 YEQSLVSYVLPKLQAIDGLTIYGPSDAESHVGVIAFNLEGLHPHDVATAMDYEGVAVRAG 362
           +E  L++YVLPKL+AI+GLTIYGPS    + G+I+FNL+ LHPHD+ATA+DYEGVAVRAG
Sbjct: 301 HEAELIAYVLPKLEAIEGLTIYGPSQPSARSGLISFNLDDLHPHDLATALDYEGVAVRAG 360

Query: 363 HHCAQPLINHLGIHSAVRASFYFYNTKEDCDKLVDAIQKTKEFFNGTL 410
           HHCAQPL+++LG+ + VRASFY YNTK DCD+LV+AI K KEFFNGTL
Sbjct: 361 HHCAQPLLSYLGVPATVRASFYIYNTKADCDRLVEAILKAKEFFNGTL 408
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1860

A DNA sequence (GBSx1967) was identified in *S.agalactiae* <SEQ ID 5781> which encodes the amino acid sequence <SEQ ID 5782>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1441(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07189 GB:AP001518 unknown conserved protein [Bacillus halodurans]
 Identities = 171/430 (39%), Positives = 267/430 (61%), Gaps = 15/430 (3%)

Query: 1   MSKEAILNFLQAKGEPTWLQELRLKAFEKIEELELPVIERVKFHRWNLG--DGTILENDY 58
           + KE + +F  A+ EP W +++RLK FE +E LELP  ++ K   WN    D  + E
Sbjct: 9   IDKEYVQSFSDARNEPQWFKDIRLKGFELVETLELPKPDKTKITSWNFTNFDHKLPEVSP 68

Query: 59  TANVPDFTE---------LGNNPKLVQIGTQTVLEQVPMELIEKGVVFTDFYSALEEIPE 109
            A++ +  +               LVQ     V  ++   L  KGV+FTD  +A++E  +
Sbjct: 69  VASIDELRDEVKGLIGEASDTQNLLVQRDATVVYSKLDEALKAKGVIFTDLLTAVKEHGD 128

Query: 110 VIERYFGK-ARPFEEDRLAAYHTAYFNSGAVLYIPDNVEITQPIEGLFYQDSQSKVPFNK 168
           ++E+Y+ K A   +E+RL A H A  N G  +Y+P NVEI  P++ +F+ D++    FN
Sbjct: 129 LVEKYYMKDAVKVDENRLTALHAALVNGGTFIYVPRNVEIEVPLQSVFWFDTEKAGLFN- 187

Query: 169 HILLIVGKNAKVSYLERFESIGDGTERTSANISVEVIAQAGSQIKFASIDRLGENVTTFI 228
           H++++   N+ ++Y+E + S G  +E   ANI VEV A A +++ F ++D L   VTT++
Sbjct: 188 HVIIVAEDNSSITYVENYASFG--SEEAVANIVVEVFAGANAKVSFGAVDNLAAGVTTYV 245

Query: 229 SRRGRHSSDATIDWALGVMNEGNVVADFDSDLIGDGSHANLKVVAASSGRQVQGIDTRVT 288
            RR     D+ ++WALG MN+GN V++  + L+GD S A+ K V+   G Q Q   T++
Sbjct: 246 VRRAHVGRDSRVEWALGQMNDGNTVSENTTHLLGDNSWADTKTVSVGRGEQKQNFTTQIF 305

Query: 289 NYGCNSVGHILQHGVILERGTLTFNGIGHIIKGAKGADAQQESRVLMLSDKARSDANPIL 348
```

```
           ++G +S G+IL+HGV+ E  T  FNGI  I  GA  +  +Q  RVLMLS+KAR DANPIL
Sbjct: 306 HHGKHSEGYILKHGVMREAATSIFNGISKIEHGATKSHGEQTERVLMLSEKARGDANPIL 365

Query: 349 LIDENDVTAGHAASIGQVDPEDLYYLMSRGLNQKTAEQLVIRGFLGTVIAEIPVKEVRDE 408
           LIDE+DVTAGHAAS+G++DP  ++YLMSRG+++  AE+LVI GFL  V+ ++P++ V++
Sbjct: 366 LIDEDDVTAGHAASVGKIDPIQMFYLMSRGISRAEAERLVIHGFLAPVVGQLPIESVKER 425

Query: 409 MIAVIDTKLE 418
           ++  I+ K++
Sbjct: 426 LVEAIERKVK 435
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5783> which encodes the amino acid sequence <SEQ ID 5784>. Analysis of this protein sequence reveals the following:

```
        Possible site: 14

   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -0.80    Transmembrane   387 - 403 ( 387 - 403)

   ----- Final Results -----
                 bacterial membrane --- Certainty=0.1319(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB15259 GB:Z99120 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 177/428 (41%), Positives = 267/428 (62%), Gaps = 15/428 (3%)

Query: 3   KEKLVAFSQAHAEPAWLQERRLAALEAIPNLELPTIERVKFHRWNLGDGT--LTENESLA 60
           +E L +FS+ H EPAWL+  RL ALE   +L +P  ++ K   WN  +      +NE L+
Sbjct: 11  QEYLKSFSEKHQEPAWLKNLRLQALEQAEDLPMPKPDKTKITNWNFTNFAKHTVDNEPLS 70

Query: 61  SVPDF-------IAIGDNPKLVQVGTQTVLEQLPMA--LIDKGVVFSDFYTALEEIPEVI 111
           S+ D        I I +  K + V        L ++  L DKGV+F+D  TA  E  +++
Sbjct: 71  SLEDLTDEVKALIDIENEDKTLYVQRDQTPAHLSLSQELKDKGVIFTDILTAAREHSDLV 130

Query: 112 EAHFGQ-ALAFDEDKLAAYHTAYFNSAAVLYVPDHLEITTPIEAIFLQDSDSDVPFNKHV 170
           E +F +  +  DE KL A H A  N  A LYVP ++++ TP++A+++ +S+    FN HV
Sbjct: 131 EKYFMKDGVKVDEHKLTALHAALVNGGAFLYVPKNVQVETPVQAVYVHESNDTALFN-HV 189

Query: 171 LVIAGKESKFTYLERFESIGNATQKISANISVEVIAQAGSQIKFSAIDRLGPSVTTYISR 230
           L++A   S  TY+E + S  N    +  NI  EVI    + + + A+D L   VTTY++R
Sbjct: 190 LIVAEDHSSVTYVENYISTVNPKDAVF-NIISEVITGDNASVTYGAVDNLSSGVTTYVNR 248

Query: 231 RGRLE-KDANIDWALAVMNEGNVIADFDSDLIGQGSQADLKVVAASSGRQVQGIDTRVTN 289
           RG    +D+ I+WAL +MN+G+ I++  ++L G G+  D K V    G Q +   T++ +
Sbjct: 249 RGAARGRDSKIEWALGLMNDGDTISENTTNLYGDGTYGDTKTVVVGRGEQTENFTTQIIH 308

Query: 290 YGQRTVGHILQHGVILERGTLTFNGIGHILKDAKGADAQQESRVLMLSDQARADANPILL 349
           +G+ + G+IL+HGV+ +  +  FNGIG I   A  A+A+QESRVLMLS++AR DANPILL
Sbjct: 309 FGKASEGYILKHGVMKDSASSIFNGIGKIEHGASKANAEQESRVLMLSEKARGDANPILL 368

Query: 350 IDENEVTAGHAASIGQVDPEDMYYLMSRGLDQETAERLVIRGFLGAVIAEIPIPSVRQEI 409
           IDE++VTAGHAAS+G+VDP  +YYLMSRG+ +E AERLVI GFL  V+ E+PI  V++++
Sbjct: 369 IDEDDVTAGHAASVGRVDPIQLYYLMSRGIPKEEAERLVIYGFLAPVVNELPIEGVKKQL 428

Query: 410 IKVLDEKL 417
           + V++ K+
Sbjct: 429 VSVIERKV 436
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 322/420 (76%), Positives = 368/420 (86%)

Query: 1   MSKEAILNFLQAKGEPTWLQELRLKAFEKIEELELPVIERVKFHRWNLGDGTILENDYTA 60
           M+KE ++ F QA  EP WLQE RL A E I  LELP IERVKFHRWNLGDGT+ EN+  A
```

```
Sbjct: 1   MTKEKLVAFSQAHAEPAWLQERRLAALEAIPNLELPTIERVKFHRWNLGDGTLTENESLA 60

Query: 61  NVPDFTELGNNPKLVQIGTQTVLEQVPMELIEKGVVFTDFYSALEEIPEVIERYFGKARP 120
           +VPDF  +G+NPKLVQ+GTQTVLEQ+PM LI+KGVVF+DFY+ALEEIPEVIE +FG+A
Sbjct: 61  SVPDFIAIGDNPKLVQVGTQTVLEQLPMALIDKGVVFSDFYTALEEIPEVIEAHFGQALA 120

Query: 121 FEEDRLAAYHTAYFNSGAVLYIPDNVEITQPIEGLFYQDSQSKVPFNKHILLIVGKNAKV 180
           F+ED+LAAYHTAYFNS AVLY+PD++EIT PIE +F QDS S VPFNKH+L+I GK +K
Sbjct: 121 FDEDKLAAYHTAYFNSAAVLYVPDHLEITTPIEAIFLQDSDSDVPFNKHVLVIAGKESKF 180

Query: 181 SYLERFESIGDGTERTSANISVEVIAQAGSQIKFASIDRLGENVTTFISRRGRHSSDATI 240
           +YLERFESIG+ T++ SANISVEVIAQAGSQIKF++IDRLG +VTT+ISRRGR   DA I
Sbjct: 181 TYLERFESIGNATQKISANISVEVIAQAGSQIKFSAIDRLGPSVTTYISRRGRLEKDANI 240

Query: 241 DWALGVMNEGNVVADFDSDLIGDGSHANLKVVAASSGRQVQGIDTRVTNYGCNSVGHILQ 300
           DWAL VMNEGNV+ADFDSDLIG GS A+LKVVAASSGRQVQGIDTRVTNYG  +VGHILQ
Sbjct: 241 DWALAVMNEGNVIADFDSDLIGQGSQADLKVVAASSGRQVQGIDTRVTNYGQRTVGHILQ 300

Query: 301 HGVILERGTLTFNGIGHIIKGAKGADAQQESRVLMLSDKARSDANPILLIDENDVTAGHA 360
           HGVILERGTLTFNGIGHI+K AKGADAQQESRVLMLSD+AR+DANPILLIDEN+VTAGHA
Sbjct: 301 HGVILERGTLTFNGIGHILKDAKGADAQQESRVLMLSDQARADANPILLIDENEVTAGHA 360

Query: 361 ASIGQVDPEDLYYLMSRGLNQKTAEQLVIRGFLGTVIAEIPVKEVRDEMIAVIDTKLEKR 420
           ASIGQVDPED+YYLMSRGL+Q+TAE+LVIRGFLG VIAEIP+  VR E+I V+D KL  R
Sbjct: 361 ASIGQVDPEDMYYLMSRGLDQETAERLVIRGFLGAVIAEIPIPSVRQEIIKVLDEKLLNR 420
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1861

A DNA sequence (GBSx1968) was identified in *S.agalactiae* <SEQ ID 5785> which encodes the amino acid sequence <SEQ ID 5786>. This protein is predicted to be ABC transporter, ATP-binding protein, Ycf16 family. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2253(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15260 GB:Z99120 similar to ABC transporter (ATP-binding
           protein) [Bacillus subtilis]
 Identities = 180/250 (72%), Positives = 212/250 (84%)

Query: 2   SVLEIKNLHVSIEDKEILKGLNLTLKTGEIAAIMGPNGTGKSTLSAAIMGNPNYEVTAGE 61
           S L IK+LHV IE KEILKG+NL +K GE  A+MGPNGTGKSTLSAAIMG+P YEVT G
Sbjct: 4   STLTIKDLHVEIEGKEILKGVNLEIKGGEFHAVMGPNGTGKSTLSAAIMGHPKYEVTKGS 63

Query: 62  ILFDGEDILELEVDERARLGLFLAMQYPSEVPGITNAEFIRAAMNAGKADDDKISIRQFI 121
           I  DG+D+LE+EVDERA+ GLFLAMQYPSE+ G+TNA+F+R+A+NA + + D+IS+ +FI
Sbjct: 64  ITLDGKDVLEMEVDERAQAGLFLAMQYPSEISGVTNADFLRSAINARREEGDEISLMKFI 123

Query: 122 TKLDEKMELLGMKEEMAERYLNEGFSGGEKKRNEILQLLMLEPKFALLDEIDSGLDIDAL 181
            K+DE ME L M  EMA+RYLNEGFSGGEKKRNEILQL+M+EPK A+LDEIDSGLDIDAL
Sbjct: 124 RKMDENMEFLEMDPEMAQRYLNEGFSGGEKKRNEILQLMMIEPKIAILDEIDSGLDIDAL 183

Query: 182 KVVSKGVNEMRGEGFGAMIITHYQRLLNYITPDKVHVMMDGKVVLSGGPELAVRLEKEGY 241
           KVVSKG+N+MR E FG ++ITHYQRLLNYITPD VHVMM G+VV SGG ELA RLE EGY
Sbjct: 184 KVVSKGINKMRSENFGCLMITHYQRLLNYITPDVVHVMMQGRVVKSGGAELAQRLEAEGY 243
```

```
Query: 242 AQIAEELGLE 251
             I +ELG+E
Sbjct: 244 DWIKQELGIE 253
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5787> which encodes the amino acid sequence <SEQ ID 5788>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2417(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 225/255 (88%), Positives = 241/255 (94%)

Query: 1   MSVLEIKNLHVSIEDKEILKGLNLTLKTGEIAAIMGPNGTGKSTLSAAIMGNPNYEVTAG 60
           MS+LEI NLHVSIE KEILKG+NLTLKTGE+AAIMGPNGTGKSTLSAAIMGNPNYEVT G
Sbjct: 1   MSILEINNLHVSIEGKEILKGVNLTLKTGEVAAIMGPNGTGKSTLSAAIMGNPNYEVTQG 60

Query: 61  EILFDGEDILELEVDERARLGLFLAMQYPSEVPGITNAEFIRAAMNAGKADDDKISIRQF 120
           +IL DG +IL+LEVDERARLGLFLAMQYPSE+PGITNAEF+RAAMNAGKAD+DKIS+R F
Sbjct: 61  QILLDGVNILDLEVDERARLGLFLAMQYPSEIPGITNAEFMRAAMNAGKADEDKISVRDF 120

Query: 121 ITKLDEKMELLGMKEEMAERYLNEGFSGGEKKRNEILQLLMLEPKFALLDEIDSGLDIDA 180
           ITKLDEKM LLGMKEEMAERYLNEGFSGGEKKRNEILQLLMLEPKFALLDEIDSGLDIDA
Sbjct: 121 ITKLDEKMALLGMKEEMAERYLNEGFSGGEKKRNEILQLLMLEPKFALLDEIDSGLDIDA 180

Query: 181 LKVVSKGVNEMRGEGFGAMIITHYQRLLNYITPDKVHVMMDGKVVLSGGPELAVRLEKEG 240
           LKVVSKGVNEMRG+ FGAMIITHYQRLLNYITPD VHVMMDG++VLSG   LA RLEKEG
Sbjct: 181 LKVVSKGVNEMRGKDFGAMIITHYQRLLNYITPDLVHVMMDGRIVLSGDAALATRLEKEG 240

Query: 241 YAQIAEELGLEYKEE 255
           YA IA++LG+EYKEE
Sbjct: 241 YAGIAQDLGIEYKEE 255
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1862

A DNA sequence (GBSx1969) was identified in *S.agalactiae* <SEQ ID 5789> which encodes the amino acid sequence <SEQ ID 5790>. This protein is predicted to be RgpG (rfe). Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.10    Transmembrane  312 - 328 ( 308 - 336)
   INTEGRAL    Likelihood =-10.03    Transmembrane   15 -  31 (   6 -  41)
   INTEGRAL    Likelihood = -9.82    Transmembrane  205 - 221 ( 197 - 226)
   INTEGRAL    Likelihood = -8.60    Transmembrane  335 - 351 ( 329 - 358)
   INTEGRAL    Likelihood = -7.48    Transmembrane  257 - 273 ( 255 - 281)
   INTEGRAL    Likelihood = -5.52    Transmembrane   60 -  76 (  56 -  79)
   INTEGRAL    Likelihood = -5.31    Transmembrane  151 - 167 ( 148 - 171)
   INTEGRAL    Likelihood = -4.88    Transmembrane   91 - 107 (  90 - 108)
   INTEGRAL    Likelihood = -4.78    Transmembrane  184 - 200 ( 177 - 203)
   INTEGRAL    Likelihood = -3.13    Transmembrane  119 - 135 ( 119 - 135)
   INTEGRAL    Likelihood = -2.97    Transmembrane  229 - 245 ( 229 - 250)

----- Final Results -----
               bacterial membrane --- Certainty=0.5840(Affirmative) < succ>
```

```
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8919> which encodes amino acid sequence <SEQ ID 8920> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:      5.18
GvH: Signal Score (-7.5): -6.19
     Possible site: 15
>>> Seems to have an uncleavable N-term signal seq
ALOM program  count: 9 value: -12.10 threshold:  0.0
   INTEGRAL    Likelihood =-12.10   Transmembrane  239 - 255 ( 235 - 263)
   INTEGRAL    Likelihood = -9.82   Transmembrane  132 - 148 ( 124 - 153)
   INTEGRAL    Likelihood = -8.60   Transmembrane  262 - 278 ( 256 - 285)
   INTEGRAL    Likelihood = -7.48   Transmembrane  184 - 200 ( 182 - 208)
   INTEGRAL    Likelihood = -5.31   Transmembrane   78 -  94 (  75 -  98)
   INTEGRAL    Likelihood = -4.88   Transmembrane   18 -  34 (  17 -  35)
   INTEGRAL    Likelihood = -4.78   Transmembrane  111 - 127 ( 104 - 130)
   INTEGRAL    Likelihood = -3.13   Transmembrane   46 -  62 (  46 -  62)
   INTEGRAL    Likelihood = -2.97   Transmembrane  156 - 172 ( 156 - 177)
   PERIPHERAL  Likelihood = 12.63     284
 modified ALOM score:   2.92

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.5840(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA82114 GB:AB022909 RgpG [Streptococcus mutans]
 Identities = 266/382 (69%), Positives = 317/382 (82%)

Query: 10  TIEYIFVLIGAFLLSIILTPIIRVISLKVGAVDKPNARRINKVPMPSSGGLAIFLSFVVT 69
            T++++ VLI   L S++LTP++R  +L+VGAVD PNARRINKVPMPS+GGLAI +SFV+
Sbjct: 7   TLKFVLVLIATLLTSLVLTPLVRFFALRVGAVDNPNARRINKVPMPSAGGLAIIISFVIA 66

Query: 70  TLFFMPMAASRHFIEVSYFHYILPVIIGGLVVTTTGFIDDIFELRPRYKMLGIIIAAIII 129
            TL  MPM         SYF YILPV++G LV+  TGFIDD++EL P+ K LGI++ A+II
Sbjct: 67  TLALMPMILKTQIGGKSYFEYILPVVLGALVIALTGFIDDVYELSPKIKFLGILLGAVII 126

Query: 130 WKFTHFRFDSFKIPIGGPLLEFGPILTFFLTVLWIISITNAINLIDGLDGLVSGVSIISL 189
            W FT FRFDSFKIP GGP+L F P L+FFLT+LW+++ITNA+NLIDGLDGLVSGVS+ISL
Sbjct: 127 WIFTDFRFDSFKIPFGGPMLHFNPFLSFFLTILWVVAITNAVNLIDGLDGLVSGVSMISL 186

Query: 190 ATMAVVSYFFLPKIDFFLTLTIVILIASIVGFFPYNYHPAIIYLGDAGALFIGFMIGVLS 249
             TM +VSYFFL   D FLTLTI +LI +I GFFPYNYHPAIIYLGD GALFIGFMI VLS
Sbjct: 187 TTMGLVSYFFLYDTDIFLTLTIFVLIFAIAGFFPYNYHPAIIYLGDTGALFIGFMISVLS 246

Query: 250 LQGLKNSTAVAVITPVIILGVPILDTAVAIVRRKLSGKKISEADKMHLHHRLLSMGFTHR 309
            LQGLKN+TAVAV+TP+I+LGVPI+DT VAI+RR LSG+K  EAD MHLHHRLL+MGFTHR
Sbjct: 247 LQGLKNATAVAVVTPIIVLGVPIVDTTVAIIRRTLSGQKFYEADNMHLHHRLLAMGFTHR 306

Query: 310 GAVLVVYGIAIIFSLIALLLNVSSRIGGIFLLLALLLAMEIFIEGLNIWGENRTPLFNLL 369
            GAVLVVYGIA+ FSL++LLLNVSSR+GGI L++ +  A+EIFIEGL IWG  RTPLF LL
Sbjct: 307 GAVLVVYGIAMFFSLVSLLLNVSSRLGGILLMIGVAFALEIFIEGLEIWGPKRTPLFRLL 366

Query: 370 KFIGNSDYRQSVIAKYSDKHQK 391
             FIGNSDYRQ V+AKY  K +K
Sbjct: 367 AFIGNSDYRQEVVAKYRRKKKK 388
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5791> which encodes the amino acid sequence <SEQ ID 5792>. Analysis of this protein sequence reveals the following:

```
     Possible site: 32
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.28    Transmembrane    9 -  25 (   1 -  33)
   INTEGRAL    Likelihood = -8.17    Transmembrane  201 - 217 ( 198 - 221)
   INTEGRAL    Likelihood = -7.64    Transmembrane  308 - 324 ( 305 - 329)
   INTEGRAL    Likelihood = -7.17    Transmembrane   55 -  71 (  51 -  74)
   INTEGRAL    Likelihood = -7.06    Transmembrane  145 - 161 ( 138 - 170)
   INTEGRAL    Likelihood = -6.58    Transmembrane  260 - 276 ( 251 - 278)
   INTEGRAL    Likelihood = -6.21    Transmembrane  180 - 196 ( 172 - 198)
   INTEGRAL    Likelihood = -5.95    Transmembrane  331 - 347 ( 330 - 353)
   INTEGRAL    Likelihood = -5.68    Transmembrane   87 - 103 (  82 - 104)
   INTEGRAL    Likelihood = -3.93    Transmembrane  113 - 129 ( 112 - 133)
   INTEGRAL    Likelihood = -2.60    Transmembrane  233 - 249 ( 232 - 250)

----- Final Results -----
               bacterial membrane --- Certainty=0.4312(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAA82114 GB:AB022909 RgpG [Streptococcus mutans]
 Identities = 289/381 (75%), Positives = 334/381 (86%), Gaps = 1/381 (0%)

Query: 5   TIDYVLVLIGALLMSLFLTPLVRFLAFRVGAVDNPNARRVNKVPMPTSGGLAIFMSFLVA 64
            T+ +VLVLI  LL SL LTPLVRF A RVGAVDNPNARR+NKVPMP++GGLAI +SF++A
Sbjct: 7   TLKFVLVLIATLLTSLVLTPLVRFFALRVGAVDNPNARRINKVPMPSAGGLAIIISFVIA 66

Query: 65  SLGLIPIASKGAMFFGQTYFSYILPVVIGATVITLTGFLDDLYELSPKLKMFGILIGAVI 124
            +L L+P+  K     G++YF YILPVV+GA VI LTGF+DD+YELSPK+K  GIL+GAVI
Sbjct: 67  TLALMPMILK-TQIGGKSYFEYILPVVLGALVIALTGFIDDVYELSPKIKFLGILLGAVI 125

Query: 125 VWAFTDFKFDSFKIPFGGPLLVFGPFLTLFLTVLWIVSITNAINLIDGLDGLVSGVSIIS 184
            +W FTDF+FDSFKIPFGGP+L F PFL+ FLT+LW+V+ITNA+NLIDGLDGLVSGVS+IS
Sbjct: 126 IWIFTDFRFDSFKIPFGGPMLHFNPFLSFFLTILWVVAITNAVNLIDGLDGLVSGVSMIS 185

Query: 185 LVTMAIVSYFFLPQKDFFLTLTILVLISAIAGFFPYNYHPAMIYLGDTGALFIGFMIGVL 244
            L TM +VSYFFL   D FLTLTI VLI AIAGFFPYNYHPA+IYLGDTGALFIGFMI VL
Sbjct: 186 LTTMGLVSYFFLYDTDIFLTLTIFVLIFAIAGFFPYNYHPAIIYLGDTGALFIGFMISVL 245

Query: 245 SLQGLKNSTAVAVVTPVIILGVPIMDTIVAIIRRSLSGQKFYEPDKMHLHHRLLSMGFTH 304
            SLQGLKN+TAVAVVTP+I+LGVPI+DT VAIIRR+LSGQKFYE D MHLHHRLL+MGFTH
Sbjct: 246 SLQGLKNATAVAVVTPIIVLGVPIVDTTVAIIRRTLSGQKFYEADNMHLHHRLLAMGFTH 305

Query: 305 RGAVLVVYGITMLFSLISLLLNVSSRIGGVLLMLGLLFGLEVFIEGLEIWGEKRTPLFNL 364
            RGAVLVVYGI M FSL+SLLLNVSSR+GG+LLM+G+ F LE+FIEGLEIWG KRTPLF L
Sbjct: 306 RGAVLVVYGIAMFFSLVSLLLNVSSRLGGILLMIGVAFALEIFIEGLEIWGPKRTPLFRL 365

Query: 365 LKFIGNSDYRQAMLLKWKEKK 385
            L FIGNSDYRQ ++ K++ KK
Sbjct: 366 LAFIGNSDYRQEVVAKYRRKK 386
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 282/384 (73%), Positives = 334/384 (86%), Gaps = 1/384 (0%)

Query: 6   MIPFTIEYIFVLIGAFLLSIILTPIIRVISLKVGAVDKPNARRINKVPMPSSGGLAIFLS 65
            M  FTI+Y+ VLIGA L+S+ LTP++R ++ +VGAVD PNARR+NKVPMP+SGGLAIF+S
Sbjct: 1   MFSFTIDYVLVLIGALLMSLFLTPLVRFLAFRVGAVDNPNARRVNKVPMPTSGGLAIFMS 60

Query: 66  FVVTTLFFMPMAAS-RHFIEVSYFHYILPVIIGGLVVTTTGFIDDIFELRPRYKMLGIII 124
            F+V +L  +P+A+    F   +YF YILPV+IG  V+T TGF+DD++EL P+ KM GI+I
Sbjct: 61  FLVASLGLIPIASKGAMFFGQTYFSYILPVVIGATVITLTGFLDDLYELSPKLKMFGILI 120

Query: 125 AAIIIWKFTHFRFDSFKIPIGGPLLEFGPILTFFLTVLWIISITNAINLIDGLDGLVSGV 184
             A+I+W FT F+FDSFKIP GGPLL FGP LT FLTVLWI+SITNAINLIDGLDGLVSGV
Sbjct: 121 GAVIVWAFTDFKFDSFKIPFGGPLLVFGPFLTLFLTVLWIVSITNAINLIDGLDGLVSGV 180
```

```
Query: 185 SIISLATMAVVSYFFLPKIDFFLTLTIVILIASIVGFFPYNYHPAIIYLGDAGALFIGFM 244
           SIISL TMA+VSYFFLP+ DFFLTLTI++LI++I GFFPYNYHPA+IYLGD GALFIGFM
Sbjct: 181 SIISLVTMAIVSYFFLPQKDFFLTLTILVLISAIAGFFPYNYHPAMIYLGDTGALFIGFM 240

Query: 245 IGVLSLQGLKNSTAVAVITPVIILGVPILDTAVAIVRRKLSGKKISEADKMHLHHRLLSM 304
           IGVLSLQGLKNSTAVAV+TPVIILGVPI+DT VAI+RR LSG+K  E DKMHLHHRLLSM
Sbjct: 241 IGVLSLQGLKNSTAVAVVTPVIILGVPIMDTIVAIIRRSLSGQKFYEPDKMHLHHRLLSM 300

Query: 305 GFTHRGAVLVVYGIAIIFSLIALLLNVSSRIGGIFLLLALLLAMEIFIEGLNIWGENRTP 364
           GFTHRGAVLVVYGI ++FSLI+LLLNVSSRIGG+ L+L LL  +E+FIEGL IWGE RTP
Sbjct: 301 GFTHRGAVLVVYGITMLFSLISLLLNVSSRIGGVLLMLGLLFGLEVFIEGLEIWGEKRTP 360

Query: 365 LFNLLKFIGNSDYRQSVIAKYSDK 388
           LFNLLKFIGNSDYRQ+++ K+ +K
Sbjct: 361 LFNLLKFIGNSDYRQAMLLKWKEK 384
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1863

A DNA sequence (GBSx1970) was identified in *S.agalactiae* <SEQ ID 5793> which encodes the amino acid sequence <SEQ ID 5794>. This protein is predicted to be negative regulator of genetic competence. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3460(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9483> which encodes amino acid sequence <SEQ ID 9484> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA82113 GB:AB022909 negative regulator of genetic competence
          [Streptococcus mutans]
 Identities = 168/248 (67%), Positives = 205/248 (81%), Gaps = 9/248 (3%)

Query: 1   MEMKQISETTLKITISMEDLEDRGMELKDFLIPQEKTEEFFYSVMDELDLPENFKNSGML 60
           MEMKQISETTLKITISMEDLE+RGMELKDFLIPQEKTEEFFY+VMDELDLPENFK SGML
Sbjct: 1   MEMKQISETTLKITISMEDLEERGMELKDFLIPQEKTEEFFYTVMDELDLPENFKGSGML 60

Query: 61  SFRVTPKKDRIDVFVTKSELSKDLNLEELADLGDISKMSPEDFFKTLEQSMLEKGDTDAH 120
           SFRVTP+ DRIDVFVTKSE++K+LNLE+L+D  DISKMSPEDFF TLE++M EKGD  A
Sbjct: 61  SFRVTPRNDRIDVFVTKSEINKNLNLEDLSDFDDISKMSPEDFFNTLEETMREKGDAAAL 120

Query: 121 AKLAEIENMMDKATQEVVEENVSEEQPEKEVETIGYVHYVFDFNIEAVVRFSQTIDFPI 180
            KLAEIE   ++ TQ+  E+  ++E+ +       YVH+V DF NI+ V+ F++T+D+ +
Sbjct: 121 DKLAEIEKREEEKTQQ--EKGETKEKRD-------YVHFVLDFPNIQQVISFAKTVDYDV 171

Query: 181 EASELYKNGKGYHMTILLDLENQPSYFANLMYARMLEHANVGTKTRAYLKEHSIQLIHDD 240
           EASEL+K    YHMT+LL+LE++P Y+A+LM+ARMLEHA  GTKTRAYL EH +QLI  D
Sbjct: 172 EASELFKESDAYHMTVLLNLEDKPDYYADLMFARMLEHAGRGTKTRAYLLEHGVQLIKAD 231

Query: 241 AISKLQMI 248
           A+ +LQMI
Sbjct: 232 ALQELQMI 239
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5795> which encodes the amino acid sequence <SEQ ID 5796>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3307(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 171/253 (67%), Positives = 209/253 (82%), Gaps = 2/253 (0%)

Query: 1   MEMKQISETTLKITISMEDLEDRGMELKDFLIPQEKTEEFFYSVMDELDLPENFKNSGML 60
           MEMKQISETTLKITISM+DLE+RGMELKDFLIPQEKTEEFFYSVMDELDLP+NFK+SGML
Sbjct: 3   MEMKQISETTLKITISMDDLEERGMELKDFLIPQEKTEEFFYSVMDELDLPDNFKDSGML 62

Query: 61  SFRVTPKKDRIDVFVTKSELSKDLNLEELADLGDISKMSPEDFFKTLEQSMLEKGDTDAH 120
           SFRVTP+KDR+DVFVTKSE++KD+NLE+LA+ GD+S+M+PEDFFK+LEQSM EKGD  AH
Sbjct: 63  SFRVTPRKDRLDVFVTKSEINKDINLEDLAEFGDMSQMTPEDFFKSLEQSMREKGDVKAH 122

Query: 121 AKLAEIENMMDKATQEVV--EENVSEEQPEKEVETIGYVHYVFDFDNIEAVVRFSQTIDF 178
            KL +IE +M+   +  +  +   ++     E E + YVHYV DF  I   V F++TIDF
Sbjct: 123 EKLEKIEEIMEDVVEATLANQSEAADPSTNHESEPLDYVHYVLDFSTITEAVAFAKTIDF 182

Query: 179 PIEASELYKNGKGYHMTILLDLENQPSYFANLMYARMLEHANVGTKTRAYLKEHSIQLIH 238
            IEASELYK    YHMTILLD++ QPSYFAN+MYAR++EHAN G+KTRAYL+EH +QL+
Sbjct: 183 SIEASELYKGSNCYHMTILLDVQQQPSYFANVMYARLIEHANPGSKTRAYLQEHGLQLML 242

Query: 239 DDAISKLQMIEMG 251
           D A+ +LQ IE+G
Sbjct: 243 DGAVEQLQKIELG 255
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1864

A DNA sequence (GBSx1971) was identified in *S.agalactiae* <SEQ ID 5797> which encodes the amino acid sequence <SEQ ID 5798>. This protein is predicted to be BacA (bacA). Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.02   Transmembrane  115 - 131 ( 111 - 135)
   INTEGRAL    Likelihood = -8.97   Transmembrane  227 - 243 ( 219 - 247)
   INTEGRAL    Likelihood = -7.86   Transmembrane   48 -  64 (  44 -  69)
   INTEGRAL    Likelihood = -7.27   Transmembrane  263 - 279 ( 260 - 279)
   INTEGRAL    Likelihood = -7.22   Transmembrane   87 - 103 (  85 - 107)
   INTEGRAL    Likelihood = -3.50   Transmembrane    2 -  18 (   1 -  19)

----- Final Results -----
               bacterial membrane --- Certainty=0.4609(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD50462 GB:AF169967 BacA [Flavobacterium johnsoniae]
 Identities = 101/275 (36%), Positives = 165/275 (59%), Gaps = 22/275 (8%)

Query: 7   LKALFLGVVEGVTEWLPVSSTGHLILVQEFMKLNQSKSFVEMFNIVIQLGAIMAVIVIYF 66
```

```
           L+A+ L V+EG+TE+LPVSSTGH+I+   F  +   + F ++F IVIQLGAI++V+V+YF
Sbjct: 4   LQAIVLAVIEGITEFLPVSSTGHMIIASSFFGIAH-EDFTKLFTIVIQLGAILSVVVLYF 62

Query: 67  KRLNPFQPGKSAREIRLTWQLWLKVVIACIPSILIALPFDNWFEAHFNFMIPIAIALIFY 126
           KR   FQ          T   + K+++A IP++++ L   ++ +      + +A++L+
Sbjct: 63  KRF--FQ----------TLDFYFKLLVAFIPAVVLGLLLSDFIDGLLENPVTVAVSLLIG 110

Query: 127 GFVFI----WVEKRNAHLKPQVTELASMSYKTAFLIGCFQVLSIVPGTSRSGATILGAII 182
           G + +    W    NA    Q      ++Y  A  IG FQ ++++PG SRSGA+I+G +
Sbjct: 111 GLILLKVDEWFNNPNAAETSQ-----KITYLQALKIGLFQCIAMIPGVSRSGASIVGGMS 165

Query: 183 IGTSRSVAADFTFFLAIPTMFGYSGLKAVKYFLDGNVLSLDQSLILLVASLTAFVVSLYV 242
              SR+ AA+F+FFLA+PTM G +  K   Y+  G  LS DQ  IL++ ++ AF+V+L
Sbjct: 166 QKLSRTTAAEFSFFLAVPTMLGATVKKCYDYYKAGFELSHDQVNILIIGNVVAFIVALLA 225

Query: 243 IRFLTDYVKRHDFTIFGKYRIVLGSLLILYWLVVH 277
           I+    ++ ++ F +FG YRI+ G +L+L    +H
Sbjct: 226 IKTFISFLTKNGFKVFGYYRIIAGIILLLIHFFIH 260
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5799> which encodes the amino acid sequence <SEQ ID 5800>. Analysis of this protein sequence reveals the following:

```
      Possible site: 56
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.30   Transmembrane  225 - 241 ( 219 - 247)
   INTEGRAL    Likelihood = -9.24   Transmembrane  115 - 131 ( 109 - 135)
   INTEGRAL    Likelihood = -7.64   Transmembrane   48 -  64 (  44 -  69)
   INTEGRAL    Likelihood = -7.43   Transmembrane   87 - 103 (  85 - 108)
   INTEGRAL    Likelihood = -5.15   Transmembrane  263 - 279 ( 262 - 279)
   INTEGRAL    Likelihood = -3.82   Transmembrane    2 -  18 (   1 -  19)

----- Final Results -----
               bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAD50462 GB:AF169967 BacA [Flavobacterium johnsoniae]
 Identities = 102/269 (37%), Positives = 169/269 (61%), Gaps = 14/269 (5%)

Query: 7   LKAIFFGIIEGITEWLPVSSTGHLILVQEFIRLNQDKAFIEMFNIVIQLGAIIAVMLIYF 66
           L+AI   +IEGITE+LPVSSTGH+I+   F  +  +  F ++F IVIQLGAI++V+++YF
Sbjct: 4   LQAIVLAVIEGITEFLPVSSTGHMIIASSFFGIAHED-FTKLFTIVIQLGAILSVVVLYF 62

Query: 67  ERLNPFQPGKTAREVQLTWQLWLKVVIACIPSILIAVPLDNWFEAHFYFMVPIAIALIVY 126
           +R   FQ          T   + K+++A IP++++ + L ++ +      V +A++L++
Sbjct: 63  KRF--FQ----------TLDFYFKLLVAFIPAVVLGLLLSDFIDGLLENPVTVAVSLLIG 110

Query: 127 GIAFIWIEKRNAQQEPAVTELARMSYKTAFFIGCFQVLSIVPGTSRSGATILGAIILGTS 186
           G+  + +++       A T   +++Y  A  IG FQ ++++PG SRSGA+I+G +    S
Sbjct: 111 GLILLKVDEWFNNPNAAETS-QKITYLQALKIGLFQCIAMIPGVSRSGASIVGGMSQKLS 169

Query: 187 RTVAADFTFFLAIPTMFGYSGLKAVKFFLDGHHLDFAQVLILLVASLTAFVVSLLAIRFL 246
           RT AA+F+FFLA+PTM G +  K   ++  G  L   QV IL++ ++ AF+V+LLAI+
Sbjct: 170 RTTAAEFSFFLAVPTMLGATVKKCYDYYKAGFELSHDQVNILIIGNVVAFIVALLAIKTF 229

Query: 247 TDYVKKHDFTIFGKYRIVLGSLLLIYSFF 275
             ++ K+ F +FG YRI+ G +LL+  FF
Sbjct: 230 ISFLTKNGFKVFGYYRIIAGIILLLIHFF 258
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 227/272 (83%), Positives = 253/272 (92%)

Query: 1   MLIIELLKALFLGVVEGVTEWLPVSSTGHLILVQEFMKLNQSKSFVEMFNIVIQLGAIMA 60
           MLIIELLKA+F G++EG+TEWLPVSSTGHLILVQEF++LNQ K+F+EMFNIVIQLGAI+A
```

```
Sbjct: 1   MLIIELLKAIFFGIIEGITEWLPVSSTGHLILVQEFIRLNQDKAFIEMFNIVIQLGAIIA 60

Query: 61  VIVIYFKRLNPFQPGKSAREIRLTWQLWLKVVIACIPSILIALPFDNWFEAHFNFMIPIA 120
           V++IYF+RLNPFQPGK+ARE++LTWQLWLKVVIACIPSILIA+P DNWFEAHF FM+PIA
Sbjct: 61  VMLIYFERLNPFQPGKTAREVQLTWQLWLKVVIACIPSILIAVPLDNWFEAHFYFMVPIA 120

Query: 121 IALIFYGFVFIWVEKRNAHLKPQVTELASMSYKTAFLIGCFQVLSIVPGTSRSGATILGA 180
           IALI YG  FIW+EKRNA  +P VTELA MSYKTAF IGCFQVLSIVPGTSRSGATILGA
Sbjct: 121 IALIVYGIAFIWIEKRNAQQEPAVTELARMSYKTAFFIGCFQVLSIVPGTSRSGATILGA 180

Query: 181 IIIGTSRSVAADFTFFLAIPTMFGYSGLKAVKYFLDGNVLSLDQSLILLVASLTAFVVSL 240
           II+GTSR+VAADFTFFLAIPTMFGYSGLKAVK+FLDG+ L   Q LILLVASLTAFVVSL
Sbjct: 181 IILGTSRTVAADFTFFLAIPTMFGYSGLKAVKFFLDGHHLDFAQVLILLVASLTAFVVSL 240

Query: 241 YVIRFLTDYVKRHDFTIFGKYRIVLGSLLILY 272
             IRFLTDYVK+HDFTIFGKYRIVLGSLL++Y
Sbjct: 241 LAIRFLTDYVKKHDFTIFGKYRIVLGSLLLIY 272
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1865

A DNA sequence (GBSx1972) was identified in *S.agalactiae* <SEQ ID 5801> which encodes the amino acid sequence <SEQ ID 5802>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.65    Transmembrane  494 - 510 ( 488 - 519)
   INTEGRAL    Likelihood = -8.01    Transmembrane  263 - 279 ( 256 - 288)
   INTEGRAL    Likelihood = -5.95    Transmembrane   25 -  41 (  20 -  43)
   INTEGRAL    Likelihood = -4.94    Transmembrane  475 - 491 ( 473 - 493)

----- Final Results -----
               bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9481> which encodes amino acid sequence <SEQ ID 9482> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB99606 GB:U67598 M. jannaschii predicted coding region MJ1577
          [Methanococcus jannaschii]
 Identities = 41/172 (23%), Positives = 78/172 (44%), Gaps = 19/172 (11%)

Query: 479 LISFVVIIYTLFLNYFTYFCIYLLLFGVILLLNKIIFMMTRKISNGYIVTEDGASRVYQW 538
           +IS ++ ++  F+  ++     +    ++ ++  II  +T     G         ++ +W
Sbjct: 442 VISILLAVFLYFIPKYSQTFNEVFYLSIVFVVQNIILALTPTSLFGRWKANYYKEKL-EW 500

Query: 539 TSFRNMLRDIKSFDRSELESIVLWNRILVYATLFGYADRVEKALR-VNQIDIPERFANID 597
            +F+N L ++    +   E I +W   L+Y T  G  D+V +A++ +N  ++   +  I
Sbjct: 501 DAFKNFLSNLAMIKKYSPEDISIWKDWLIYGTALGVGDKVVEAMKSLNLSELVADYVIIH 560

Query: 598 SHQFAISVNQSSNHFSTITEDVSHASNFSVNSGGSSGGFSGGGG--GGGGGA 647
           S+  ++  +  S + ST                GS GGF  GGG  GGGGGA
Sbjct: 561 SNYDSMKTSVDSVYSSTT---------------GSGGGFGAGGGFGGGGGGA 597
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5803> which encodes the amino acid sequence <SEQ ID 5804>. Analysis of this protein sequence reveals the following:

```
     Possible site: 21
>>> Seems to have a cleavable N-term signal seq.
```

```
  INTEGRAL    Likelihood = -7.91    Transmembrane  486 - 502 ( 483 - 508)
  INTEGRAL    Likelihood = -5.89    Transmembrane  465 - 481 ( 460 - 483)
  INTEGRAL    Likelihood = -2.18    Transmembrane  244 - 260 ( 241 - 260)

----- Final Results -----
              bacterial membrane --- Certainty=0.4163(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAB99606 GB:U67598 M. jannaschii predicted coding region MJ1577
           [Methanococcus jannaschii]
 Identities = 59/263 (22%), Positives = 106/263 (39%), Gaps = 14/263 (5%)

Query: 369 FLDMAFGNKVTLPVDQLFSQYHYDADTIKQLKKTYKGKKLEQEVRQSSEQVIKAMKKASA 428
           ++ +  G K+ +    L +   Y++D +K L K  K    + E  +S  Q  K+ K
Sbjct: 346 YIKIMNGGKIEILKTDLENLDVYESDVMKFLMKYSKNNVFDPEYIKSLAQKYKSSKDKLK 405

Query: 429 AITNNVLETIKKLNLPDTYRQMTPA--EKRKSNSVQGLGCLLLILNSGLLIYLAIKESGL 486
            + +   E  K +  P    ++  A  E R    +  L  + ++L   L      ++
Sbjct: 406 KLKD---ELDKIMEYPRYSSKVVNAFLETRGKKIIIALLVISILLAVFLYFIPKYSQTFN 462

Query: 487 ALIYLALMVLTMCLGFYISLKLDQYKKLGIETPEGGVRLHQWQSFKNMIRDIDKFEDVAI 546
            + YL+++ +       I L L      G           +W +FKN + ++   +  +
Sbjct: 463 EVFYLSIVFVVQ----NIILALTPTSLFGRWKANYYKEKLEWDAFKNFLSNLAMIKKYSP 518

Query: 547 EGLVVWNRVLVYATLFGYAKKVERYLKVHRIALPEVYQAVRPGELSMVMYATTPTFVSSL 606
           E + +W   L+Y T  G   KV   +K   ++     + V    +    Y +  T V S+
Sbjct: 519 EDISIWKDWLIYGTALGVGDKVVEAMKSLNLS-----ELVADYVIIHSNYDSMKTSVDSV 573

Query: 607 SSATTSSNFSVSSGGGISGGGGG 629
            S+TT S     +GGG  GGGGG
Sbjct: 574 YSSTTGSGGGFGAGGGFGGGGGG 596
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 241/635 (37%), Positives = 372/635 (57%), Gaps = 18/635 (2%)

Query: 22  MKKCFLAICLALSFFMVSVQADEVDYNIPHYEGNLTIHNDNSADFTEKVTYQFDSSYNGQ 81
           MKK  + + L  S   + ++A +VDY+I +YEG L +  +N+A F +KVTYQFD+SYNGQ
Sbjct: 1   MKKILMTLVLCFSLLGIRIKAADVDYSITNYEGQLLLSKENTARFEQKVTYQFDTSYNGQ 60

Query: 82  YVTLGTAGKLPDNFDINNKPQVEVSINGKVRKVSYQIEDLEDGYRLKVFNGGEAGDTVKV 141
           Y++LG  G LP  F I+ KP+VEV  NG+   VS +  DL DGYRLK++N G+AGD V V
Sbjct: 61  YISLGRTGHLPAGFAIDQKPKVEVYQNGQQVPVSQEFSDLGDGYRLKLYNAGQAGDKVDV 120

Query: 142 NVQWKLKNVLFMHKDVGELNWIPISDWDKTLEKVDFWISTDKKVALSRLWGHLGYL-KTP 200
            V W+L ++L  ++DV ELNW PISDWDKTLEKV   ++T   +  S LW H GY  K P
Sbjct: 121 KVIWQLHHLLTAYQDVAELNWTPISDWDKTLEKVSLTVTTPTDIQDSNLWAHRGYYQKKP 180

Query: 201 PKIRQNNNRYHLTAFNVNKRLEFHGYWDRSYF--NLPTNSKNNYKKKIEYQEKMIERHGF 258
             +++ N+RY + A NV+ +LE H YWD+       P +    + K KI   E  I R
Sbjct: 181 QVLKEGNSRYQINAKNVSGQLELHAYWDKKALLGKEPVDVSTSKKNKIVALETKISRRRT 240

Query: 259 ILSFLLRILLPSFFIIVTLFISIRVFLFRKKVNKYGQFPKEHHLYEAPEDLSPLELTQSI 318
           +L  L   ++P   +   L+  I+    +K+ N+Y       H YE PEDLSPL LTQ+I
Sbjct: 241 LLQLLFGKVIPLVEVGFLLWQLIQFTRLKKQFNRYHLANHTDHSYEVPEDLSPLVLTQAI 300

Query: 319 YSMSFKNFQ---DEEKKTHL---ISQEQLIQSILLDLIDRKVL----NYDDNLLSLANLD 368
           Y  SF        E +K  +   ++ E L+Q+ LLDLID+KVL          L ++ LD
Sbjct: 301 YGQSFAYLSPTASESQKLLIPKGVTFEALVQATLLDLIDQKVLLLTKEEGKAYLEISQLD 360

Query: 369 RASDAEIDFIEFAFADSTSLKPDQLFSNYQFSYKETLRELKKQHKASDLQTQMRRRGSNA 428
           R +D E  F++ AF +  +L  DQLFS Y +   +T+++LKK +K   L+ ++R+
Sbjct: 361 RVTDEEAAFLDMAFGNKVTLPVDQLFSQYHYD-ADTIKQLKKTYKGKKLEQEVRQSSEQV 419

Query: 429 LSRITRLTRLISKDNINSLRRKGISSPYRKMSSEESKELSRLKRFSYLSPLISFVVIIYT 488
```

```
           +  + + +  I+ + + ++++  +   YR+M+  E ++ + ++    L  +++  ++IY
Sbjct: 420 IKAMKKASAAITNNVLETIKKLNLPDTYRQMTPAEKRKSNSVQGLGCLLLILNSGLLIY- 478

Query: 489 LFLNYFTYFCIYLLLFGVILLLNKIIFMMTRKISNGYIVTEDGASRVYQWTSFRNMLRDI 548
           L +       IYL L  + + L   I +   +     I T +G  R++QW SF+NM+RDI
Sbjct: 479 LAIKESGLALIYLALMVLTMCLGFYISLKLDQYKKLGIETPEGGVRLHQWQSFKNMIRDI 538

Query: 549 KSFDRSELESIVLWNRILVYATLFGYADRVEKALRVNQIDIPERFANIDSHQFAISVNQS 608
             F+   +E +V+WNR+LVYATLFGYA +VE+ L+V++I +PE +  +   + ++ +  +
Sbjct: 539 DKFEDVAIEGLVVWNRVLVYATLFGYAKKVERYLKVHRIALPEVYQAVRPGELSMVMYAT 598

Query: 609 SNHFSTITEDVSHASNFSVNSGGSSGGFSGGGGGG 643
           +  F +     + +SNFSV+SG   GG SGGGGGG
Sbjct: 599 TPTFVSSLSSATTSSNFSVSSG---GGISGGGGGG 630
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8921> and protein <SEQ ID 8922> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 4
McG: Discrim Score:     10.29
GvH: Signal Score (-7.5): 3.11
     Possible site: 23
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 3 value:  -8.65 threshold:  0.0
   INTEGRAL    Likelihood = -8.65   Transmembrane  475 - 491 ( 469 - 500)
   INTEGRAL    Likelihood = -8.01   Transmembrane  244 - 260 ( 237 - 269)
   INTEGRAL    Likelihood = -4.94   Transmembrane  456 - 472 ( 454 - 474)
   PERIPHERAL  Likelihood =  2.28     540
 modified ALOM score:   2.23

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no homology with any sequences in the databases.

Example 1866

A DNA sequence (GBSx1973) was identified in *S.agalactiae* <SEQ ID 5805> which encodes the amino acid sequence <SEQ ID 5806>. This protein is predicted to be glutamine-binding periplasmic protein/glutamine transport system perme. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -8.86   Transmembrane  301 - 317 ( 295 - 324)
   INTEGRAL    Likelihood = -6.05   Transmembrane  479 - 495 ( 473 - 496)
   INTEGRAL    Likelihood = -0.59   Transmembrane  369 - 385 ( 369 - 385)

----- Final Results -----
               bacterial membrane --- Certainty=0.4545(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA17584 GB:D90907 glutamine-binding periplasmic protein
          [Synechocystis sp.]
 Identities = 147/534 (27%), Positives = 256/534 (47%), Gaps = 75/534 (14%)
```

```
Query: 4   ILLSLFTALLITFGGMTSIQADEYLRVGMEAAYAPFNWTQNDNTNGAVPIEGTDQYANGY 63
           +LL++   LL  F  ++     + + V  E  + PF  T           E T Q   G+
Sbjct: 24  VLLAIAIPLLPAFSQVSR----QTIIVATEPTFPPFEMTD----------EATGQLT-GF 68

Query: 64  DVQVAKKLAKKLNKKVVVVKTKWEGLVPALTSGKLDMIIAGMSPTEERKKEINFSKPYYI 123
           DV + + + +     V +    ++G++PAL S  +   I+ ++ T ER + ++FS PY+
Sbjct: 69  DVDLIQAIGEAAQVTVDIQGYPFDGIIPALQSNTVGAAISAITITPERAQSVSFSSPYFK 128

Query: 124 SEPTLVVNAEGKYTNAKNISDFKNAKVTAQQGVYLYNLIDQINGVKKEVAMGDFNQLRQA 183
           S   L +  +      KN+ D +  ++    G     +   + G K    + +F+ +  A
Sbjct: 129 S--VLAIAVQDGNDTIKNLKDLEGKRLAVAIGTTGAMVATNVPGAK----VTNFDSITSA 182

Query: 184 VE---SGVVDAYVSERPDATSAQTANPKLKMIELHQGFKTSDADTNISVGMRKGDNRINQ 240
           ++   +G  DA +++RP    A   +  L+ +++     + D    I++ +      INQ
Sbjct: 183 LQELVNGNADAVINDRPVLLYA-IKDAGLRNVKISADVGSEDY-YGIAMPLAP-PGEINQ 239

Query: 241 VNQVL-----ESISRDKQIALMDKMIKEQ---------PSV------------KKEKNGK 274
             +VL     + I      A+ +K   E+         PS+            + + N
Sbjct: 240 TREVLNQGLFQIIENGTYNAIYEKWFGEKNPPFLPLVAPSLVGKVGTAQSLTERSQANPN 299

Query: 275 PNFFEQMATILKNNGSQFLRGTATTLLISMVGTIVGLFIGLLIGVFRTAPKSDNKLKAAL 334
            NF   + T+ +N      +G+  T+L++      GL  G  + +   A  SD
Sbjct: 300 DNF---LITLFRN----LFKGSILTVLLTAFSVFFGLIGGTGVAI---ALISD------- 342

Query: 335 QKLLGWLLNIYIEVFRGTPMIVQSMVIYYGTAQAF-----GVSLDRTLAAIFIVSINTGA 389
            K L  +  IY+E FRGTPM+VQ  +IY+G    F     G+++DR  AAI  +S+N  A
Sbjct: 343 IKPLQLIFRIYVEFFRGTPMLVQLFIIYFGLPALFKEIGLGITIDRFPAAIIALSLNVAA 402

Query: 390 YMSEIVRGGIFSVDKGQFEAATALGFTHGQTMRKIVLPQVVRNILPATGNEFVINIKDTS 449
           Y++EI+RGGI S+D+GQ+EA  +LG +  QTM++++ PQ  R ILP  GNEF+  IKDTS
Sbjct: 403 YLAEIIRGGIQSIDQGQWEACESLGMSPWQTMKEVIFPQAFRRILPPLGNEFITLIKDTS 462

Query: 450 VLNVISVVELYFSGNTVATQTYQYFQTFTIIAIIYFILTFTVTRILRYIEKRFD 503
           +  VI   EL+  G  +   TY+ F+ +  +A++Y +LT   + + +++E   D
Sbjct: 463 LTAVIGFQELFREGQLIVATTYRAFEVYIAVALVYLLLTTISSFVFKWLENYMD 516
```

There is also homology to SEQ ID 1194.

A related GBS gene <SEQ ID 8923> and protein <SEQ ID 8924> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 2
McG: Discrim Score:      6.23
GvH: Signal Score (-7.5): 0.11
     Possible site: 24
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 3 value:  -8.86 threshold:  0.0
   INTEGRAL    Likelihood = -8.86   Transmembrane  301 - 317 ( 295 - 324)
   INTEGRAL    Likelihood = -6.05   Transmembrane  479 - 495 ( 473 - 496)
   PERIPHERAL  Likelihood =  1.32     441
 modified ALOM score:   2.27

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.4545(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
34.3/57.3% over 462aa
                                                           Synechocystis PCC6803
  EGAD|48193| glutamine-binding periplasmic protein/glutamine transport system permease
protein Insert characterized
```

GP|1652664|dbj|BAA17584.1||D90907 glutamine-binding periplasmic protein {Synechocystis sp.} Insert characterized

PIR|S77250|S77250 hypothetical protein - Synechocystis sp. (strain PCC 6803) Insert characterized

ORF01242(454 - 1809 of 2148)
EGAD|48193|sll1270(54 - 516 of 530) glutamine-binding periplasmic protein/glutamine transport system permease protein {Synechocystis PCC6803}GP|1652664|dbj|BAA17584.1||D90907 glutamine-binding periplasmic protein {Synechocystis sp.}PIR|S77250|S77250 hypothetical protein - Synechocystis sp. (strain PCC 6803)
%Match = 12.3
%Identity = 34.2 %Similarity = 57.2
Matches = 128 Mismatches = 149 Conservative Sub.s = 86

```
204       234       264       294       324       354       384       414
PSFVCIPF*HKNTINRFQ*DNDIEIDLVFR*NRRK*LIGGC*MKKILLSLFTALLITXGGMTSIQADEYLRVGMEAAYAP

                                        MKGMVKLGHWGKTWRYYLLLALGVLLAIAIPLLPAFSQVS
                                                10        20        30        40

444       474       495       525       555       585       615       645
FNWTQNDNTNGAVPIEGTDQ---YANGYDVQVAKKLAKKLNKKVVVVKTKWEGLVPALTSGKLDMIIAGMSPTEERKKEI
             | | ||:      |:|| : : : :     | :    ::|::||| |  :   |: :: | || : :
RQTIIVATEPTFPPFEMTDEATGQLTGFDVDLIQAIGEAAQVTVDIQGYPFDGIIPALQSNTVGAAISAITITPERAQSV
        50        60        70        80        90       100       110       120

675       705       735       765       795       825       855       885
NFSKPYYISEPTLVVNAEGKYTNAKNISDFKNAKVTAQQGVYLYNLIDQINGVKKEVAMGDFNQLRQAVESGVVDAYVSE
:|| ||: |   : |  :|  |  ||: |::  ::    |     :   : | |        : |:: | :|  || :::
SFSSPYFKSVLAIAVQ-DGNDT-IKNLKDLEGKRLAVAIGTTGAMVATNVPGAKVTNFDSITSALQELV-NGNADAVIND
       130        140        150       160       170       180        190

903                 957       987
RP--------------------~~~~--------------------DATSAQTANPKLK-MIELHQG-FKTSDADTNISV
||                                              |    |||    :| | :  || |
RPVLLYAIKDAGLRNVKISADV~~~~NPPFLPLVAPSLVGKVGTAQSLTERSQANPNDNFLITLFRNLFKGS--------
          210            270       280       290       300       310

1017      1047      1077      1107      1137      1167      1197      1227
GMRKGDNRINQVNQVLESISRDKQIALMDKMIKEQPSVKKEKNGKPNFFEQMATILKNNGSQFLRGTATTLLISMVGTIV
                                                                       ::::: |
-----------------------------------------------------------------------ILTVLLTAF
                                                                          320

1257      1284      1314      1344      1374      1404      1419      1449
GLFIGLLIGV-FRTAPKSDNKLKAALQKLLGWLLNIYIEVFRGTPMIVQSMVIYYGTAQAF-----GVSLDRTLAAIFIV
 :| ||: |     |  ||        | |  :: ||:| ||||||:||  :||:|    |     |:::||  |||  :
SVFFGLIGGTGVAIALISD-------IKPLQLIFRIYVEFFRGTPMLVQLFIIYFGLPALFKEIGLGITIDRFPAAIIAL
              340              350       360       370       380       390

1479      1509      1539      1569      1599      1629      1659      1689
SINTGAYMSEIVRGGIFSVDKGQFEAATALGFTHGQTMRKIVLPQVVRNILPATGNEFVINIKDTSVLNVISVVELYFSG
|:|  ||::||:|||| |:|:||:||  :|| :  |||:::: ||  | |||  ||||:  |||||:  ||   ||:  |
SLNVAAYLAEIIRGGIQSIDQGQWEACESLGMSPWQTMKEVIFPQAFRRILPPLGNEFITLIKDTSLTAVIGFQELFREG
           410       420       430       440       450       460       470

1719      1749      1779      1809      1839      1869      1899      1929
NTVATQTYQYFQTFTIIAIIYFILTFTVTRILRYIEKRFDSDNYTTGANQLQV*EVGMTQAILEIKHLKKSYGSNEVLKD
  :   ||: |: :  :|::|::||   : :::::|   |         :
QLIVATTYRAFEVYIAVALVYLLLTTISSFVFKWLENYMDPIGRAKKKAKAATA
           490       500       510       520       530
```

There is also homology to SEQ ID 5804.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1867

A DNA sequence (GBSx1974) was identified in *S.agalactiae* <SEQ ID 5807> which encodes the amino acid sequence <SEQ ID 5808>. This protein is predicted to be ATP-binding. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3208(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB73160 GB:AL139076 putative glutamine transport ATP-binding
          protein [Campylobacter jejuni]
 Identities = 132/241 (54%), Positives = 178/241 (73%), Gaps = 1/241 (0%)

Query: 5   ILEIKHLKKSYGSNEVLKDISLSVNKGEVISIIGSSGSGKSTFLRSINLLEEPSGGEILY 64
           ++E+K+L+K YG  EVLK+I+ +++KG+VI+IIG SG GKSTFLR IN LE    GEIL
Sbjct: 1   MIEVKNLQKKYGELEVLKNINTTISKGDVIAIIGPSGGGKSTFLRCINRLELADSGEILI 60

Query: 65  HGHNVLEKGYDLNNYREKLGMVFQSFNLFENLNILENAIVAQTTVLKRERQEAEKIAKEN 124
           +  N+L+K  D+N  R+K+ MVFQ FNLF N N++EN  +         ++EA K AK
Sbjct: 61  NKQNILDKEIDINKIRQKVSMVFQHFNLFANKNVMENLCLTPIKTGILSQEEAIKKAKLL 120

Query: 125 LNAVGMTEQYWKAKPKQLSGGQKQRVAIARALSVNPEAILFDEPTSALDPEMVGEVLKTM 184
           L  VG+ ++     P +LSGGQKQR+AIAR+L +NP+ ILFDEPTSALDPEM+GEVL  M
Sbjct: 121 LAKVGLADKE-NIMPHKLSGGQKQRIAIARSLMMNPDVILFDEPTSALDPEMIGEVLSIM 179

Query: 185 QDLAKSGLTMIIVTHEMEFAKEVSDRVIFMDKGIIAEQGTPKQLFENPTQERTKEFLQRFL 245
           +D+AK GLTM++VTHEM FA+ V++R+ FMDKG IA   +PK++FENP+ ER +EFL + L
Sbjct: 180 KDVAKEGLTMLVVTHEMGFARNVANRIFFMDKGKIAVDASPKEVFENPSNERLREFLNKVL 240
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2157> which encodes the amino acid sequence <SEQ ID 2158>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1170(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 212/246 (86%), Positives = 237/246 (96%)

Query: 1   MTQAILEIKHLKKSYGSNEVLKDISLSVNKGEVISIIGSSGSGKSTFLRSINLLEEPSGG 60
           M+ +I+EIK+LKKSYGSNEVLKDISLSVNKGEVISIIGSSGSGKST LRSINLLEEPS G
Sbjct: 24  MSNSIIEIKNLKKSYGSNEVLKDISLSVNKGEVISIIGSSGSGKSTLLRSINLLEEPSAG 83

Query: 61  EILYHGHNVLEKGYDLNNYREKLGMVFQSFNLFENLNILENAIVAQTTVLKRERQEAEKI 120
           +IL+HG +VL + Y+L +YREKLGMVFQSFNLFENLN+LENAIVAQTTVLKR+R +AE+I
Sbjct: 84  QILFHGEDVLAEHYNLTHYREKLGMVFQSFNLFENLNVLENAIVAQTTVLKRDRAQAEQI 143

Query: 121 AKENLNAVGMTEQYWKAKPKQLSGGQKQRVAIARALSVNPEAILFDEPTSALDPEMVGEV 180
           AKENLNAVGMTEQYW+AKPKQLSGGQKQRVAIARALSVNPEA+LFDEPTSALDPEMVGEV
Sbjct: 144 AKENLNAVGMTEQYWQAKPKQLSGGQKQRVAIARALSVNPEAMLFDEPTSALDPEMVGEV 203

Query: 181 LKTMQDLAKSGLTMIIVTHEMEFAKEVSDRVIFMDKGIIAEQGTPKQLFENPTQERTKEF 240
           LKTMQDLAKSGLTMIIVTHEMEFA++VSDR+IFMDKG+I E+G+P+Q+FENPTQ+RTKEF
Sbjct: 204 LKTMQDLAKSGLTMIIVTHEMEFARDVSDRIIFMDKGLITEEGSPQQIFENPTQDRTKEF 263
```

```
Query: 241 LQRFLK 246
           LQRFLK
Sbjct: 264 LQRFLK 269
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1868

A DNA sequence (GBSx1976) was identified in *S.agalactiae* <SEQ ID 5809> which encodes the amino acid sequence <SEQ ID 5810>. This protein is predicted to be hypersensitive-induced response protein. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-17.94    Transmembrane    4 -  20 (    1 -  28)

----- Final Results -----
               bacterial membrane --- Certainty=0.8175(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9479> which encodes amino acid sequence <SEQ ID 9480> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF68390 GB:AF236374 hypersensitive-induced response protein
          [Zea mays]
 Identities = 127/275 (46%), Positives = 174/275 (63%), Gaps = 1/275 (0%)

Query: 19  ITSLYVVKQQTVAIIERFGKYQKTATSGIHIRVPLGIDKIAARVQLRLLQSEIIVETKTK 78
            I  L  V Q TVAI E FGK+ +    G H        +IA  + LR+ Q ++  ETKTK
Sbjct: 4   ILGLVQVDQSTVAIKENFGKFSEVLEPGCHFLPWCIGQQIAGYLSLRVRQLDVRCETKTK 63

Query: 79  DNVFVTLNIATQYRVNENNVTDAYYKLIKPEAQIKSYIEDALRSSVPKLTLDELFEKKDE 138
            DNVFVT+  + QYR   +  +DA+YKL     QI+SY+ D +R++VPKL LD+ FE+K+E
Sbjct: 64  DNVFVTVVASVQYRALADKASDAFYKLSNTREQIQSYVFDVIRATVPKLGLDDAFEQKNE 123

Query: 139 IALEVQHQVAEEMSTYGYIIVKTLITKVEPDAEVKQSMNEINAAQRKRVAAQELANADKI 198
            IA  V+ ++ + MSTYGY IV+TLI  +EPD  VK++MNEINAA R RVAA E A A+KI
Sbjct: 124 IAKAVEEELEKAMSTYGYQIVQTLIVDIEPDDRVKRAMNEINAAARMRVAASEKAEAEKI 183

Query: 199 KIVTAAEAEAEKDRLHGVGIAQQRKAIVDGLADSIQELKDANVTLTEEQIMSILLTNQYL 258
              +  AE EAE   L GVGIA+QR+AIVDGL DS+    +     T + IM ++L  QY
Sbjct: 184 LQIKKAEGEAESKYLAGVGIARQRQAIVDGLRDSVLAFSENVPGTTAKDIMDMVLVTQYF 243

Query: 259 DTLNTF-AINGNQTIFLPNNPEGVEDIRTQVLSAL 292
            DT+    A + + ++F+P+ P  V+D+  Q+   L
Sbjct: 244 DTMREIGASSKSSSVFIPHGPGAVKDVSAQIRDGL 278
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5811> which encodes the amino acid sequence <SEQ ID 5812>. Analysis of this protein sequence reveals the following:

```
     Possible site: 32
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-13.06    Transmembrane    5 -  21 (    1 -  29)

----- Final Results -----
               bacterial membrane --- Certainty=0.6222(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF68390 GB:AF236374 hypersensitive-induced response protein
         [Zea mays]
 Identities = 126/273 (46%), Positives = 174/273 (63%), Gaps = 3/273 (1%)

Query: 23  LYVVRQQSVAIVERFGRYQKTATSGIHIRLPFGI-DKIAARVQLRLLQSEIIVETKTKDN 81
           L  V Q +VAI E FG++ +    G H  LP+ I  +IA  + LR+ Q ++  ETKTKDN
Sbjct: 7   LVQVDQSTVAIKENFGKFSEVLEPGCHF-LPWCIGQQIAGYLSLRVRQLDVRCETKTKDN 65

Query: 82  VFVTLNVATQYRVNEQNVTDAYYKLMKPESQIKSYIEDALRSSVPKLTLDELFEKKDEIA 141
           VFVT+  + QYR      +DA+YKL     QI+SY+ D +R++VPKL LD+ FE+K+EIA
Sbjct: 66  VFVTVVASVQYRALADKASDAFYKLSNTREQIQSYVFDVIRATVPKLGLDDAFEQKNEIA 125

Query: 142 LEVQHQVAEEMSTYGYIIVKTLITKVEPDAEVKQSMNEINAAQRKRVAAQELANADKIKI 201
             V+ ++ + MSTYGY IV+TLI  +EPD  VK++MNEINAA R RVAA E A A+KI
Sbjct: 126 KAVEEELEKAMSTYGYQIVQTLIVDIEPDDRVKRAMNEINAAARMRVAASEKAEAEKILQ 185

Query: 202 VTAAEAEAEKDRLHGVGIAQQRKAIVDGLAESIQELKEANISLNEEQIMSILLTNQYLDT 261
           +  AE EAE   L GVGIA+QR+AIVDGL +S+    E       + IM ++L  QY DT
Sbjct: 186 IKKAEGEAESKYLAGVGIARQRQAIVDGLRDSVLAFSENVPGTTAKDIMDMVLVTQYFDT 245

Query: 262 LNTFAAKG-NQTLFLPNTPSGVEDIRTQVLSAL 293
           +    A   + ++F+P+ P  V+D+  Q+   L
Sbjct: 246 MREIGASSKSSSVFIPHGPGAVKDVSAQIRDGL 278
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 254/291 (87%), Positives = 278/291 (95%)

Query: 5   IILTVILVLVIVLLITSLYVVKQQTVAIIERFGKYQKTATSGIHIRVPLGIDKIAARVQL 64
           I +   +++++ ++ ++LYVV+QQ+VAI+ERFG+YQKTATSGIHIR+P GIDKIAARVQL
Sbjct: 6   IFIAFGVIVILAIVASTLYVVRQQSVAIVERFGRYQKTATSGIHIRLPFGIDKIAARVQL 65

Query: 65  RLLQSEIIVETKTKDNVFVTLNIATQYRVNENNVTDAYYKLIKPEAQIKSYIEDALRSSV 124
           RLLQSEIIVETKTKDNVFVTLN+ATQYRVNE NVTDAYYKL+KPE+QIKSYIEDALRSSV
Sbjct: 66  RLLQSEIIVETKTKDNVFVTLNVATQYRVNEQNVTDAYYKLMKPESQIKSYIEDALRSSV 125

Query: 125 PKLTLDELFEKKDEIALEVQHQVAEEMSTYGYIIVKTLITKVEPDAEVKQSMNEINAAQR 184
           PKLTLDELFEKKDEIALEVQHQVAEEMSTYGYIIVKTLITKVEPDAEVKQSMNEINAAQR
Sbjct: 126 PKLTLDELFEKKDEIALEVQHQVAEEMSTYGYIIVKTLITKVEPDAEVKQSMNEINAAQR 185

Query: 185 KRVAAQELANADKIKIVTAAEAEAEKDRLHGVGIAQQRKAIVDGLADSIQELKDANVTLT 244
           KRVAAQELANADKIKIVTAAEAEAEKDRLHGVGIAQQRKAIVDGLA+SIQELK+AN++L
Sbjct: 186 KRVAAQELANADKIKIVTAAEAEAEKDRLHGVGIAQQRKAIVDGLAESIQELKEANISLN 245

Query: 245 EEQIMSILLTNQYLDTLNTFAINGNQTIFLPNNPEGVEDIRTQVLSALKTR 295
           EEQIMSILLTNQYLDTLNTFA  GNQT+FLPN P GVEDIRTQVLSALKT+
Sbjct: 246 EEQIMSILLTNQYLDTLNTFAAKGNQTLFLPNTPSGVEDIRTQVLSALKTK 296
```

SEQ ID 5810 (GBS231) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 55 (lane 7; MW 60.9kDa).

GBS231d was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 155 (lane 5-7; MW 59kDa) and in Figure 239 (lane 11; MW 59kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 155 (lane 9; MW 34kDa) and in Figure 183 (lane 6; MW 34kDa). Purified GBS231d-GST is shown in Figure 246, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1869

A DNA sequence (GBSx1977) was identified in *S.agalactiae* <SEQ ID 5813> which encodes the amino acid sequence <SEQ ID 5814>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2305(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9291> which encodes amino acid sequence <SEQ ID 9292> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB13457 GB:Z99112 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 259/514 (50%), Positives = 350/514 (67%), Gaps = 9/514 (1%)

Query: 1   MGMTMENGAKEVSDKPATTVGEVGQILSKGVLMGARGNSGVITSQLFRGFGQSIKDKEEL 60
           M ++M +GA+EV       +G+VG  LSKG+LMGARGNSGVI SQLFRGF ++I+ K+E+
Sbjct: 46  MNLSMTSGAREVEQMDTDDIGKVGSALSKGLLMGARGNSGVILSQLFRGFSKNIETKKEI 105

Query: 61  TGQDLAHAFQNGVEVAYKAVMKPVEGTILTVSRGAATAALKKAEETDDAVEVMRATLKGA 120
              + A A Q GV++AYKAVMKPVEGTILTV++ AA  A+  AE+  D   +M A  + A
Sbjct: 106 NALEFAAALQAGVDMAYKAVMKPVEGTILTVAKDAAKKAMILAEKETDITALMTAVTEEA 165

Query: 121 KRALAKTPDMLPVLKEVGVVDSGGQGLVFIYEGFLSALTGEYIASEDFKATPATMTEMVN 180
           + +L +TP++LPVLKEVGVVDSGG+GL+ +YEGFL++L GE +     KA   ++ +MV+
Sbjct: 166 EASLNRTPELLPVLKEVGVVDSGGKGLLCVYEGFLASLKGETVPQ---KAVLPSLDDMVS 222

Query: 181 AEHHKAVVGHVATEDIKYGYCTEVMVGLKQGPTYVKEFNYEEFQGYLSNLGDSLLVVNDD 240
           AEHHK+    + TEDI++G+CTEVMV L Q     +EF+   F+  LS  GDSLLV+ D+
Sbjct: 223 AEHHKSAQSMMNTEDIEFGFCTEVMVRLDQTK---REFDEGTFRQDLSQFGDSLLVIADE 279

Query: 241 EIVKVHVHTEDPGLVMQEGLKYGSLVKVKVENMRNQHDA---QMQKVEVEETVKETKEYG 297
            + KVH+H E+PG V+     YG L+K+K+ENMR QH +   Q  K    ET    + YG
Sbjct: 280 SLAKVHIHAEEPGNVLNYAQHYGELIKIKIENMREQHTSIISQESKPADNETPPAKQPYG 339

Query: 298 IIAVVAGDGLAEIFKSQGVDYIISGGQTMNPSTEDIVKAIEKVNARNVIILPNNKNIFMA 357
           I+ V  G+G+A++FKS G   +I GGQTMNPSTEDIV A++ VNA  V ILPNN NI MA
Sbjct: 340 IVTVAMGEGIADLFKSIGASVVIEGGQTMNPSTEDIVDAVKSVNADTVFILPNNSNIIMA 399

Query: 358 AQSAADVVDIPAAVVETRTVPQGFTSLLAFDPAKSLETNVADMTNSLSDVISGSVTLAVR 417
           A  AA VVD    V+  +TVPQG ++LLAF+P +  E N A+M +++  V SG VT +VR
Sbjct: 400 ANQAASVVDEQVFVIPAKTVPQGMSALLAFNPDQEAEANEANMLSAIQQVKSGQVTFSVR 459

Query: 418 DTTIDGLEIHENDILGMVDGKILVSTPDMEKALKDTFDKMIDEDSEIVTIYVGEDGKQAL 477
           DT IDG +I + D +G+++G I+ ++ +   A K    +MI ED EIVTI  GED  Q
Sbjct: 460 DTHIDGKDIKKGDFMGILNGTIIGTSENQLSAAKMLLSEMIGEDDEIVTILYGEDASQEE 519

Query: 478 AETLSEYLEETYEDVEVEIHQGDQPVYPYLMSVE 511
           AE L  +L E YE++EVEIH G QP+Y Y++S E
Sbjct: 520 AEQLEAFLSEKYEEIEVEIHNGKQPLYSYIVSAE 553
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5635> which encodes the amino acid sequence <SEQ ID 5636>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1816(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 434/511 (84%), Positives = 475/511 (92%)

Query: 1   MGMTMENGAKEVSDKPATTVGEVGQILSKGVLMGARGNSGVITSQLFRGFGQSIKDKEEL 60
           M MTM+NGAKEV+DKPA+TVGEVGQ+LSKG+LMGARGNSGVITSQLFRGFGQSIK K+EL
Sbjct: 44  MSMTMDNGAKEVADKPASTVGEVGQMLSKGLLMGARGNSGVITSQLFRGFGQSIKGKDEL 103

Query: 61  TGQDLAHAFQNGVEVAYKAVMKPVEGTILTVSRGAATAALKKAEETDDAVEVMRATLKGA 120
           TG+DLA AFQ GVEVAYKAVMKPVEGTILTVSRGAATAALKKA+ TDDAVEVM+A L GA
Sbjct: 104 TGKDLAQAFQVGVEVAYKAVMKPVEGTILTVSRGAATAALKKADLTDDAVEVMQAALDGA 163

Query: 121 KRALAKTPDMLPVLKEVGVVDSGGQGLVFIYEGFLSALTGEYIASEDFKATPATMTEMVN 180
           K ALAKTPD+LPVLKEVGVVDSGGQGLVFIYEGFLSAL G+Y+ S DFKATPA M+EM+N
Sbjct: 164 KGALAKTPDLLPVLKEVGVVDSGGQGLVFIYEGFLSALNGDYVTSADFKATPANMSEMIN 223

Query: 181 AEHHKAVVGHVATEDIKYGYCTEVMVGLKQGPTYVKEFNYEEFQGYLSNLGDSLLVVNDD 240
           AEHHK+VVGHVATEDI YGYCTE+MV LKQGPTYVKEFNY+EFQGYLS LGDSLLVVNDD
Sbjct: 224 AEHHKSVVGHVATEDITYGYCTEIMVALKQGPTYVKEFNYDEFQGYLSGLGDSLLVVNDD 283

Query: 241 EIVKVHVHTEDPGLVMQEGLKYGSLVKVKVENMRNQHDAQMQKVEVEETVKETKEYGIIA 300
           EIVKVHVHTEDPGLVMQEGLKYGSL+K+KV+NMRNQH+AQ+QK +VE+   E K++G+IA
Sbjct: 284 EIVKVHVHTEDPGLVMQEGLKYGSLIKIKVDNMRNQHEAQVQKTDVEKNKAEVKDFGLIA 343

Query: 301 VVAGDGLAEIFKSQGVDYIISGGQTMNPSTEDIVKAIEKVNARNVIILPNNKNIFMAAQS 360
           VVAG+GL+EIFK+QGVDY+ISGGQTMNPSTEDIVKAIE VNA+ VIILPNNKNIFMAAQS
Sbjct: 344 VVAGEGLSEIFKAQGVDYVISGGQTMNPSTEDIVKAIEAVNAKQVIILPNNKNIFMAAQS 403

Query: 361 AADVVDIPAAVVETRTVPQGFTSLLAFDPAKSLETNVADMTNSLSDVISGSVTLAVRDTT 420
           AA+VVDIPAAVV TRTVPQGFTSLLAFDP+KSLE NVADM+ SLSDV+SGSVTLAVRDTT
Sbjct: 404 AAEVVDIPAAVVATRTVPQGFTSLLAFDPSKSLEDNVADMSTSLSDVVSGSVTLAVRDTT 463

Query: 421 IDGLEIHENDILGMVDGKILVSTPDMEKALKDTFDKMIDEDSEIVTIYVGEDGKQALAET 480
           IDGLEIHEND LGMVDGKI+VS PDME  LK  F+KMIDEDSEIVTI+VGE+G Q LAE
Sbjct: 464 IDGLEIHENDFLGMVDGKIIVSNPDMEATLKAAFEKMIDEDSEIVTIFVGEEGDQDLAEE 523

Query: 481 LSEYLEETYEDVEVEIHQGDQPVYPYLMSVE 511
           L+ YL ETYEDVEVEIHQGDQPVYPYLMSVE
Sbjct: 524 LAGYLGETYEDVEVEIHQGDQPVYPYLMSVE 554
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1870

A DNA sequence (GBSx1978) was identified in *S.agalactiae* <SEQ ID 5815> which encodes the amino acid sequence <SEQ ID 5816>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.4771(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1871

A DNA sequence (GBSx1979) was identified in *S.agalactiae* <SEQ ID 5817> which encodes the amino acid sequence <SEQ ID 5818>. This protein is predicted to be proliferating-cell nucleolar antigen P120. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3774(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9345> which encodes amino acid sequence <SEQ ID 9346> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC74905 GB:AE000278 putative nucleolar proteins [Escherichia
           coli K12]
 Identities = 87/229 (37%), Positives = 128/229 (54%), Gaps = 8/229 (3%)

Query: 63   GKSIEHTTGLVYSQEPAAQ--IVAQIAEPQEGMKVLDLAAAPGGKTTHLLSYLNNTGLLV 120
            G + EH +GL Y QE ++   + A  A+     +V+D+AAAPG KTT + + +NN G ++
Sbjct: 89   GSTAEHLSGLFYIQEASSMLPVAALFADGNAPQRVMDVAAAPGSKTTQISARMNNEGAIL 148

Query: 121  SNEISNKRSKILVENVERFGARNVIVTNESSQRLAKCFNSFFDLIVFDGPCSGEGMFRKD 180
            +NE S  R K+L  N+ R G  NV +T+   +         FD I+ D PCSGEG+ RKD
Sbjct: 149  ANEFSASRVKVLHANISRCGISNVALTHFDGRVFGAAVPEMFDAILLDAPCSGEGVVRKD 208

Query: 181  PQAIQYWHKDYPTECAQLQRDILKEAIKMLAHGGILVYSTCTWSPEENEEVVNWLLQEY- 239
            P A++ W  +   E A  QR+++  A   L  GG LVYSTCT + EENE V  WL + Y
Sbjct: 209  PDALKNWSPESNQEIAATQRELIDSAFHALRPGGTLVYSTCTLNQEENEAVCLWLKETYP 268

Query: 240  ---DYLELVDIPKLNGMVEGINVPQVARMYPHHFQGEGQFVAKLRDTRS 285
               ++L L D+    G  + +       ++P  +  EG FVA+LR T++
Sbjct: 269  DAVEFLPLGDL--FPGANKALTEEGFLHVFPQIYDCEGFFVARLRKTQA 315
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5819> which encodes the amino acid sequence <SEQ ID 5820>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2316(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 213/311 (68%), Positives = 254/311 (81%), Gaps = 3/311 (0%)

Query: 1    MKLPNEFIEKYQTILKDEAEAFFDSFEQKPISAYRTNPLKEKQLDFPNAIPSTPWGHYGK 60
            M LP EFI  YQ IL  E E F  SF Q+P++A+R NPLK +   F + IP+T WG+YGK
Sbjct: 2    MSLPKEFINTYQAILGKELEDFLASFNQEPVNAFRINPLKNQLKTFEHPIPNTLWGYYGK 61

Query: 61   ISGKSIEHTTGLVYSQEPAAQIVAQIAEPQEGMKVLDLAAAPGGKTTHLLSYLNNTGLLV 120
            +SGKS EH +GLVYSQEPAAQ+VAQ+A PQ+G +VLDLAAAPGGK+THLL+YL+NTGLLV
Sbjct: 62   LSGKSPEHVSGLVYSQEPAAQMVAQVAAPQKGSRVLDLAAAPGGKSTHLLAYLDNTGLLV 121

Query: 121  SNEISNKRSKILVENVERFGARNVIVTNESSQRLAKCFNSFFDLIVFDGPCSGEGMFRKD 180
            SNEIS KRSK+LVEN+ERFGARNV+VTNES+ RLAK F+ +FD IVFDGPCSGEGMFRKD
Sbjct: 122  SNEISKKRSKVLVENIERFGARNVVVTNESADRLAKVFSHYFDTIVFDGPCSGEGMFRKD 181
```

```
Query: 181 PQAIQYWHKDYPTECAQLQRDILKEAIKMLAHGGILVYSTCTWSPEENEEVVNWLLQEYD 240
           P AIQYWH  YP ECA+LQ+ IL++A+ ML  GG L+YSTCTW+PEENE+VV WLL+ Y
Sbjct: 182 PDAIQYWHHGYPAECAKLQKSILEDALAMLKPGGELIYSTCTWAPEENEDVVQWLLETYT 241

Query: 241 YLELVDIPKLNGMVEGINVPQVARMYPHHFQGEGQFVAKLRDTRSKEAQKIKPKAQKIN- 299
           +LELVD+PKLNGMV GI +P+ ARMYPH +QGEGQFVAKL+D R +E Q  K KA K N
Sbjct: 242 FLELVDVPKLNGMVSGIGLPETARMYPHRYQGEGQFVAKLKDKR-QEGQSTKLKAPKSNL 300

Query: 300 -KMQLQLWQQF 309
            K QL+LW+ F
Sbjct: 301 IKDQLRLWKMF 311
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1872

A DNA sequence (GBSx1980) was identified in *S.agalactiae* <SEQ ID 5821> which encodes the amino acid sequence <SEQ ID 5822>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4111(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC24940 GB:AF012285 unknown [Bacillus subtilis]
 Identities = 86/240 (35%), Positives = 133/240 (54%), Gaps = 10/240 (4%)

Query: 6   DFAKQLVYKAGQFIKSEMQNTFDVEEKSRFDDLVTSLDKKTQKLLIQEIIQHYPDDNILA 65
           + AK+ + +AG  I   M  +  +E KS  +DLVT++DK+T+K  I  I + +P   IL
Sbjct: 9   EIAKKWIREAGARITQSMHESLTIETKSNPNDLVTNIDKETEKFFIDRIQETFPGHRILG 68

Query: 66  EE---DBVRSPIAQGNVWVLDPIDGTVNFIVQKDNFAVMLAYYEEGVGQFGIIYDVMADI 122
           EE   D + S   +G VW++DPIDGT+NF+ Q+ NFA+ +  +E G G+ G+IYDV+ D
Sbjct: 69  EEGQGDKIHS--LEGVVWIIDPIDGTMNFVHQQRNFAISIGIFENGEGKIGLIYDVVHDE 126

Query: 123 LYSGGGHFDVYANDKKIVPFQECPLERCLLGVNSAMYAEN----DCGIAHLASETLGVRI 178
           LY        Y N+ K+ P +E  +E  +L +N+    EN       +A L     G R
Sbjct: 127 LYHAFSGRGAYMNETKLAPLKETVIEEAILAINATWVTENRRIDQSVLAPLVKRVRGTRS 186

Query: 179 YGGAGISMAKVMQGKLLAYFSY-IQPWDYAAAKIMGETLGFTLLTLDGEEPNYSTRQKVM 237
           YG A + +A V  G++ AY +  + PWDYAA  ++   +G T  T++GE   +     V+
Sbjct: 187 YGSAALELANVAAGRIDAYITMRLAPWDYAAGCVLLNEVGGTYTTIEGEPFTFLENHSVL 246
```

A related GBS nucleic acid sequence <SEQ ID 10937> which encodes amino acid sequence <SEQ ID 10938> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5823> which encodes the amino acid sequence <SEQ ID 5824>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1843(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 155/253 (61%), Positives = 205/253 (80%)

Query: 1   MDAKFDFAKQLVYKAGQFIKSEMQNTFDVEEKSRFDDLVTSLDKKTQKLLIQEIIQHYPD 60
            ++ K+ FA+Q++ +AG FIKS+M    D++ K++FDDLVT++D++TQ+LL+  I Q YP
Sbjct: 8   LETKYAFARQIIKEAGLFIKSKMSEQLDIQVKTQFDDLVTNVDQETQQLLMDRIHQTYPC 67

Query: 61  DNILAEEDBVRSPIAQGNVWVLDPIDGTVNFIVQKDNFAVMLAYYEEGVGQFGIIYDVMA 120
            D ILAEE++VR PI QGNVWV+DPIDGTVNFIVQ   FAVM+AYYE+G+GQFG+IYDVMA
Sbjct: 68  DAILAEENDVRHPINQGNVWVIDPIDGTVNFIVQGSQFAVMIAYYEQGIGQFGLIYDVMA 127

Query: 121 DILYSGGGHFDVYANDKKIVPFQECPLERCLLGVNSAMYAENDCGIAHLASETLGVRIYG 180
            D L +GGG F+V  N  K+  +QE PLER L+G N+ M+A ND  +AHL ++TLGVR+YG
Sbjct: 128 DQLLAGGGDFEVTLNGDKLPAYQEKPLERSLIGCNAGMFARNDRNLAHLIAKTLGVRVYG 187

Query: 181 GAGISMAKVMQGKLLAYFSYIQPWDYAAAKIMGETLGFTLLTLDGEEPNYSTRQKVMFLP 240
            GAGI M KVM+ +LLAYFS+IQPWDYAAAK++G+ LG+ LLT+DG EP++ TRQK+MF+P
Sbjct: 188 GAGICMVKVMKQELLAYFSFIQPWDYAAAKVLGDKLGYVLLTIDGYEPDFQTRQKIMFVP 247

Query: 241 KSKLNLIQSYLTK 253
            K +L  I S+LTK
Sbjct: 248 KCQLTRIASFLTK 260
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1873

A DNA sequence (GBSx1981) was identified in *S.agalactiae* <SEQ ID 5825> which encodes the amino acid sequence <SEQ ID 5826>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4131(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC24938 GB:AF012285 unknown [Bacillus subtilis]
 Identities = 33/78 (42%), Positives = 50/78 (63%)

Query: 13 YSYPLDPSWNTEDITKVLRFLNQVEHAYENSIKVDDLLDSYKEFKKVVKSKAQEKQIDRE 72
          Y YP++  W TE+   V+ F  QVE AYE     ++LL +Y+ FK++V  KA+EK++  E
Sbjct: 3  YQYPMNEDWTTEEAVDVIAFFQQVELAYEKGADREELLKAYRRFKEIVPGKAEEKKLCGE 62

Query: 73 FQRTSGYSTYQAVKAAQQ 90
          F+  S YS Y+ VK A++
Sbjct: 63 FEEQSTYSPYRTVKQARE 80
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5827> which encodes the amino acid sequence <SEQ ID 5828>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4442(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 59/91 (64%), Positives = 70/91 (76%)

Query: 9  ISSNYSYPLDPSWNTEDITKVLRFLNQVEHAYENSIKVDDLLDSYKEFKKVVKSKAQEKQ 68
           +S NY YPLD SW+TE+I+ VL FLN+VE AYE  +    LLDSYK +K +VKSKAQEKQ
Sbjct: 5  MSGNYYYPLDLSWSTEEISSVLHFLNKVELAYEKKVDAKQLLDSYKTYKTIVKSKAQEKQ 64

Query: 69 IDREFQRTSGYSTYQAVKAAQQQAKGFISLG 99
           IDR+FQ+ SGYSTYQ VK A+   KGF SLG
Sbjct: 65 IDRDFQKVSGYSTYQVVKKAKAIEKGFFSLG 95
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1874

A DNA sequence (GBSx1982) was identified in *S.agalactiae* <SEQ ID 5829> which encodes the amino acid sequence <SEQ ID 5830>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence (or aa 1-18)

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0952(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF21893 GB:AF103794 unknown [Listeria monocytogenes]
 Identities = 74/126 (58%), Positives = 101/126 (79%)

Query: 1   MITLFLSPSCTSCRKARAWLSKHEVAFEEHNIITSPLNKEELLQILSFTENGTEDIISTR 60
            M+TL+ SPSCTSCRK+RAWL +H++ ++E NI + PL+ +E+ +IL  TE+GT++IISTR
Sbjct: 1   MVTLYTSPSCTSCRKSRAWLEEHDIPYKERNIFSEPLSLDEIKEILRMTEDGTDEIISTR 60

Query: 61  SKVFQKLAIDVDELSTSSLMELISENPSLLRRPIILDKKRMQIGFNEDEIRAFLPRDYRK 120
            SK FQKL +D+D L     L ELI +NP LLRRPII+D+KR+Q+G+NEDEIR FLPR  R
Sbjct: 61  SKTFQKLNVDLDSLPLQQLFELIQKNPGLLRRPIIIDEKRLQVGYNEDEIRRFLPRRVRT 120

Query: 121 QELKQA 126
             +L++A
Sbjct: 121 YQLREA 126
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5831> which encodes the amino acid sequence <SEQ ID 5832>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0511(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 112/134 (83%), Positives = 127/134 (94%)

Query: 1   MITLFLSPSCTSCRKARAWLSKHEVAFEEHNIITSPLNKEELLQILSFTENGTEDIISTR 60
            M+TLFLSPSCTSCRKARAWL KHEV F+EHNIITSPL+++EL+ ILSFTENGTEDIISTR
Sbjct: 1   MVTLFLSPSCTSCRKARAWLVKHEVDFQEHNIITSPLSRDELMSILSFTENGTEDIISTR 60

Query: 61  SKVFQKLAIDVDELSTSSLMELISENPSLLRRPIILDKKRMQIGFNEDEIRAFLPRDYRK 120
            SKVFQKL IDV+ELS S L++LI++NPSLLRRPII+D+KRMQIGFNEDEIRAFL RDYRK
Sbjct: 61  SKVFQKLDIDVEELSISDLIDLIAKNPSLLRRPIIMDQKRMQIGFNEDEIRAFLSRDYRK 120
```

```
Query: 121 QELKQATIRAEIEG 134
           QEL+QATI+AEIEG
Sbjct: 121 QELRQATIKAEIEG 134
```

SEQ ID 5830 (GBS232) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 10; MW 16.8kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 56 (lane 2; MW 42kDa).

GBS232-GST was purified as shown in Figure 207, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1875

A DNA sequence (GBSx1983) was identified in *S.agalactiae* <SEQ ID 5833> which encodes the amino acid sequence <SEQ ID 5834>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5835> which encodes the amino acid sequence <SEQ ID 5836>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1768(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 210/308 (68%), Positives = 252/308 (81%)

Query: 1   MKIHYINDYKDIQAKEDCVLVLGYFDGLHLGHKALFDKAKKIATEKNLKIVVLTFNETPR 60
           M+I YI DY+DI  ++D VL+LGYFDGLH GHKALFDKA+++A ++ LK+VV TF E+P+
Sbjct: 1   MEIEYIKDYRDINQEDDTVLILGYFDGLHRGHKALFDKAREVANKEGLKVVVFTFTESPK 60

Query: 61  LTFARFQPELLLHLTSPEKRSEKFQEYGVDELYLMNFTSHFSKVSSDLFIKKYIYGLRAK 120
           L F+RF PELLLH+T P+KR EKF +YGV++LYL++FTS FSKVSSD FI  YI  L+AK
Sbjct: 61  LAFSRFSPELLLHITYPKKRYEKFADYGVNKLYLVDFTSKFSKVSSDHFITHYIKNLKAK 120

Query: 121 AAVVGFDYKFGHNRTSGDYLARNFKGPVYIIDEISEGGEKISSTRIRQLITEGNVEKANQ 180
             VVGFDYKFGHNRT  DYL RNF+G VY I+EI E   KIS+T IR+LI EGNV KAN
Sbjct: 121 HIVVGFDYKFGHNRTDSDYLTRNFEGQVYTIEEIKEDHRKISATWIRKLIQEGNVVKANH 180

Query: 181 LLGYEFSTCGMVVHGDARGRTIGFPTANLAPINRTYLPADGVYISNVLINGKYYRAMTSI 240
           LLGY+ ST G VVHGDARGRTIGFPTANLAPI+ TYLPADGVY++NV++  K YR+MTS+
Sbjct: 181 LLGYDLSTRGRVVHGDARGRTIGFPTANLAPIDNTYLPADGVYVTNVIVANKIYRSMTSL 240

Query: 241 GKNITFGGTELRLEANIFDFDGDIYGETIEIFWLKRIREMVKFNGIDDLVKQLKKDKEIA 300
           GKN+TFGG ELRLE NIFDFD +IYGE IEI WL +IR+M KF GI+DL  +L+ DK  A
Sbjct: 241 GKNVTFGGKELRLEVNIFDFDEEIYGEIIEIVWLDKIRDMEKFEGIEDLTDRLEYDKRTA 300

Query: 301 LNWKKDSQ 308
```

```
           LNWKKDS+
Sbjct: 301 LNWKKDSK 308
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1876

A DNA sequence (GBSx1984) was identified in *S.agalactiae* <SEQ ID 5837> which encodes the amino acid sequence <SEQ ID 5838>. This protein is predicted to be tRNA pseudouridine 5S synthase (truB). Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2576(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9817> which encodes amino acid sequence <SEQ ID 9818> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06129 GB:AP001515 tRNA pseudouridine 5S synthase [Bacillus halodurans]
 Identities = 145/283 (51%), Positives = 191/283 (67%), Gaps = 12/283 (4%)

Query: 2   ITGIINLKKEAGMTSHDAVFKLRKILHTKKIGHGGTLDPDVVGVLPIAVGKATRVIEYMT 61
           +TGI+ L K  GMTSHD V KLR++L TKK+GH GTLDPDV GVLP+ +G AT+V +YM+
Sbjct: 3   MTGILPLAKPRGMTSHDCVAKLRRLLKTKKVGHTGTLDPDVYGVLPVCIGHATKVAQYMS 62

Query: 62  ESGKIYEGEITLGYATSTEDSSGEVISRTPLTQSDLSEDVVDHAMKSFTGPITQVPPMYS 121
           +  K YEGE+T+G++T+TED SG+ +  T   Q    E VVD  + +F G I Q+PPMYS
Sbjct: 63  DYPKAYEGEVTVGFSTTTEDRSGDTVE-TKTIQQPFVEAVVDQVLATFVGEIKQIPPMYS 121

Query: 122 AVKVNGKKLYEYARSGEEVERPKRQITISEFRRTSPLYFEKGICRFSFYVSCSKGTYVRT 181
           AVKV GK+LYEYAR+G  VERP+R +TI    R S + +E+G+CRF F VSCSKGTYVRT
Sbjct: 122 AVKVRGKRLYEYARAGITVERPERTVTIFSLERMSDIVYEEGVCRFRFNVSCSKGTYVRT 181

Query: 182 LAVDLGIKLGYASHMSFLKRTSSAGLSITQSLTLEEINEKYKQ-EDFSFLLPIEYGVLDL 240
           LAVD+G  LGY +HMS L RT S   S+ +  T  E+ E+ +Q E  S LLPIE  +LD+
Sbjct: 182 LAVDIGKALGYPAHMSDLVRTKSGPFSLEECFTFTELEERLEQGEGSSLLLPIETAILDI 241

Query: 241 PKVNLTEEDKVEISYGR----------RILLENEADTLAAFYE 273
           P+V + +E + +I +G           R  + NE   L A Y+
Sbjct: 242 PRVQVNKEIEEKIRHGAVLPQKWFNHPRFTVYNEEGALLAIYK 284
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5839> which encodes the amino acid sequence <SEQ ID 5840>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2698(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 201/295 (68%), Positives = 246/295 (83%), Gaps = 2/295 (0%)
```

```
Query: 1   MITGIINLKKEAGMTSHDAVFKLRKILHTKKIGHGGTLDPDVVGVLPIAVGKATRVIEYM 60
           MI GIINLKKEAGMTSHDAVFKLRK+L  KKIGHGGTLDPDVVGVLPIAVGKATRVIEYM
Sbjct: 1   MINGIINLKKEAGMTSHDAVFKLRKLLQEKKIGHGGTLDPDVVGVLPIAVGKATRVIEYM 60

Query: 61  TESGKIYEGEITLGYATSTEDSSGEVISRTPLTQSDLSEDVVDHAMKSFTGPITQVPPMY 120
           TE+GK+YEG++TLGY+T+TED+SGEV++R+ L  + L+E++VD  M +F G ITQ PPMY
Sbjct: 61  TEAGKVYEGQVTLGYSTTTEDASGEVVARSSL-PAVLTEELVDQTMTTFLGKITQTPPMY 119

Query: 121 SAVKVNGKKLYEYARSGEEVERPKRQITISEFRRTSPLYF-EKGICRFSFYVSCSKGTYV 179
           SAVKVNG+KLYEYAR+GE VERP+R++TIS F RTSPL F E G+CRFSF V+CSKGTYV
Sbjct: 120 SAVKVNGRKLYEYARAGESVERPRREVTISLFERTSPLNFTEDGLCRFSFKVACSKGTYV 179

Query: 180 RTLAVDLGIKLGYASHMSFLKRTSSAGLSITQSLTLEEINEKYKQEDFSFLLPIEYGVLD 239
           RTLAVDLG  LG  SHMSFL+R++SAGL++  + TL EI +   +++ SFLLPIEYGV D
Sbjct: 180 RTLAVDLGRALGVESHMSFLQRSASAGLTLETAYTLGEIADMVSKQEMSFLLPIEYGVAD 239

Query: 240 LPKVNLTEEDKVEISYGRRILLENEADTLAAFYENRVIAILEKRGNEFKPHKVLL 294
           LPK+ + + +  EIS+GRR+ L ++   LAAF+  +VIAILEKR  E+KP KVL+
Sbjct: 240 LPKMVIDDTELTEISFGRRLSLPSQEPLLAAFHGEKVIAILEKRDQEYKPKKVLI 294
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1877

A DNA sequence (GBSx1985) was identified in *S.agalactiae* <SEQ ID 5841> which encodes the amino acid sequence <SEQ ID 5842>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2776(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9819> which encodes amino acid sequence <SEQ ID 9820> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12871 GB:Z99109 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 39/145 (26%), Positives = 68/145 (46%), Gaps = 7/145 (4%)

Query: 3   MKIRTATLDDSEKLVPLYQELG----YAISLSEIQSILKVILTHSDYGFLIAEDNGKLLA 58
           M IR A   D+  + PL+ +       A  L   ++ LK  L + +   LIAE+NG+ +
Sbjct: 1   MNIRQAKTSDAAAIAPLFNQYREFYRQASDLQGAEAFLKARLENHESVILIAEENGEFIG 60

Query: 59  FVGYHKLYFFEKSGTYYRILALVVNEKHRRKGIASQLINHVKQLAKTDGSEVLALNSSLK 118
           F   +  +        Y +  L V    R KG   +L++  K  A  +G++ L L +  +
Sbjct: 61  FTQLYPTFSSVSMKRIYILNDLFVVPHARTKGAGGRLLSAAKDYAGQNGAKCLTLQT--E 118

Query: 119 EYRQEAYHFYENLGFKKVSTGFSYY 143
            + ++A   YE  G+++  TGF +Y
Sbjct: 119 HHNRKARSLYEQNGYEE-DTGFVHY 142
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5843> which encodes the amino acid sequence <SEQ ID 5844>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0962(Affirmative) < succ>
```

```
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 37/126 (29%), Positives = 64/126 (50%), Gaps = 16/126 (12%)

Query: 18  PLYQE-----LGYAISLSEIQSILKVILTHSDYGFLIA--EDNGKLLAFVG---YHKLYF 67
            P+ QE     LGY +SL  ++   + ++    + FL    +D  +LL +V    Y  LY
Sbjct: 11  PMLQEINAKALGYLVSLDLLERQYERLIEDCHHYFLAYADKDTNQLLGYVHAERYETLY- 69

Query: 68  FEKSGTYYRILALVVNEKHRRKGIASQLINHVKQLAKTDGSEVLALNSSLKEYRQEAYHF 127
               +     +L L V   ++R+GI S L+  ++  A+ +G   + LNS+   +R+EA+ F
Sbjct: 70  ---ASDGLNLLGLAVLPAYQRRGIGSALLRALESQARQEGIAFIRLNSA--SHRKEAHAF 124

Query: 128 YENLGF 133
            Y NL +
Sbjct: 125 YRNLDY 130
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1878

A DNA sequence (GBSx1986) was identified in *S.agalactiae* <SEQ ID 5845> which encodes the amino acid sequence <SEQ ID 5846>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1659(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>

RGD motif 28-30
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF30776 GB:AE002133 conserved hypothetical [Ureaplasma
           urealyticum]
 Identities = 106/440 (24%), Positives = 206/440 (46%), Gaps = 65/440 (14%)

Query: 13  FAINESEYHQLLEQIRGDAFDKEVSERLEKERLILGEQAKNQLQEVVVE-KDKEIAKLQY 71
            F  N+ +Y++L++Q     +D      LEK+R  L E+ KN+  + +   KD +  K
Sbjct: 71  FLANDRDYNELVKQ----RYD------LEKQRDELKEKLKNEGNKAIAHFKDSDEYKNLI 120

Query: 72  KVKQFLIEKDNLLKDNEYQLAEQLNQKDMMLRD--------LENQIDRLRLEHENSLQEA 123
            K ++ +   +  ++ NE    +++   ++ L+         L+N I +  ++ +N+ + A
Sbjct: 121 KAQEKINSLNKTIESNEQSYKKEIENIELKLKSQFDEETKSLKNTIAKQEIKLDNAEKMA 180

Query: 124 LTKVERE-------RDAIQNQLHIQ--------------------EKEKDLALASVKSDY 156
            +   +         +D I   + I+                    E +K + +  ++S
Sbjct: 181 IINFKESNEYQKIIKDKIDLDIEIEKLKFAIQAHEDNMKAAKENWESKKIVEIKELESKK 240

Query: 157 EVQLKAANEQVEFYKNFKAQQSTKAVGESLEHYAETEFNKVRHLAFPNAYFEKDNTLSSR 216
            + ++    E +E  K  K+  + K VGE LE + + +F++    + P+  F K N
Sbjct: 241 DKEIHKLTESIEQLKREKSS-NVKLVGEELEQWLKNKFDETYSFSCPDMTFTKINEAID- 298

Query: 217 GSKGDFIY------REKDENDLEFL-SIMFEMKNESDDTIKKHKNEDFFKELDKDRREKS 269
            G K DF+       +E   +D + + S   E K E  D  K  KN   +K+LD+DR  +
Sbjct: 299 GKKADFLLEFFDFGKEMSNDDKKLIFSATIEAKTEFFDNQKGTKNSAHYKKLDQDRINQK 358

Query: 270 CEYAVLVTMLEADNDYYNTGIVDVSHKYPKMYVIRPQFFIQLIGILRNAALNTLKYKQEL 329
             EYA+LVT LE ++ +    ++   ++Y  M+ +RPQ+FI L+ ++RN A  TLK K
Sbjct: 359 SEYAILVTELEPEDHF----VIKKINEYKNMFAVRPQYFIPLVDMIRNFA--TLKAKINS 412
```

```
Query: 330 ALMKEQNIDITHFEEDLDIFKNAFAKN-YNSASKNFQKAIDEIDKSIKRMEAV-KAALTT 387
            +++ +  D    EE+LD  K     N     +   +K ID+    IK+ E++ ++A
Sbjct: 413 QIIRYE--DRAKIEENLDELKKDIVDNTLKYINDKTKKIIDDSKAIIKKAESIEESAEDI 470

Query: 388 SENQLRLANNKLDDVSVKKL 407
              +L     K+++++++K+
Sbjct: 471 INKKLNTLKKKINELTIRKI 490
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5847> which encodes the amino acid sequence <SEQ ID 5848>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3192(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 310/445 (69%), Positives = 352/445 (78%), Gaps = 22/445 (4%)

Query: 1   MNEIKCPHCGTAFAINESEYHQLLEQIRGDAFDKEVSERLEKERLILGEQAKNQLQEVVV 60
           MNEIKCPHC T F INESEY QLLEQ+RG AFD+E+ +RL  E  +L E+AK+QL EVV
Sbjct: 1   MNEIKCPHCHTLFTINESEYSQLLEQVRGQAFDEELKKRLINEIALLEEKAKHQLHEVVA 60

Query: 61  EKDKEIAKLQYKVKQF-----------LIEKDNLL-----------KDNEYQLAEQLNQK 98
           +K+  I  L  +++Q            L +KD L+             N  +LA QL +K
Sbjct: 61  KKETAITSLTNQLEQIEKEQAYLRQEELAKKDQLIASLEAKLDKLASQNALELANQLAEK 120

Query: 99  DMMLRDLENQIDRLRLEHENSLQEALTKVERERDAIQNQLHIQEKEKDLALASVKSDYEV 158
           D  +  L NQ+D+L LE + + Q  L  +E+ERD I+NQL +Q KE +L+LASV+SDYE
Sbjct: 121 DKEVVSLTNQLDKLALEKDATFQSKLATIEKERDGIKNQLALQAKESELSLASVRSDYEA 180

Query: 159 QLKAANEQVEFYKNFKAQQSTKAVGESLEHYAETEFNKVRHLAFPNAYFEKDNTLSSRGS 218
           QLKAANEQVEFYKNFKAQQSTKA+GESLE YAETEFNKVR  AFPNA F KDN LSSRGS
Sbjct: 181 QLKAANEQVEFYKNFKAQQSTKAIGESLELYAETEFNKVRSYAFPNASFVKDNQLSSRGS 240

Query: 219 KGDFIYREKDENDLEFLSIMFEMKNESDDTIKKHKNEDFFKELDKDRREKSCEYAVLVTM 278
           KGD+IYRE D N +E LSIMFEMKNE+D T  KHKN DFFKELDKDRREK CEYAVLV+M
Sbjct: 241 KGDYIYREVDANGVEILSIMFEMKNEADTTKTKHKNSDFFKELDKDRREKDCEYAVLVSM 300

Query: 279 LEADNDYYNTGIVDVSHKYPKMYVIRPQFFIQLIGILRNAALNTLKYKQELALMKEQNID 338
           LEADNDYYNTGIVDVSH+Y KMYV+RPQ FIQLIGILRNAALN+L YKQELAL+KEQNID
Sbjct: 301 LEADNDYYNTGIVDVSHEYQKMYVVRPQLFIQLIGILRNAALNSLHYKQELALVKEQNID 360

Query: 339 ITHFEEDLDIFKNAFAKNYNSASKNFQKAIDEIDKSIKRMEAVKAALTTSENQLRLANNK 398
           ITHFEEDLD FKNAFAKNY SAS NF+KAIDEIDKSIKRME VK  LTTSENQLRLANNK
Sbjct: 361 ITHFEEDLDQFKNAFAKNYQSASNNFKKAIDEIDKSIKRMEEVKRFLTTSENQLRLANNK 420

Query: 399 LDDVSVKKLTRKNPTMKAKFDALKD 423
           L+DVSVKKLTR+NPTM+ KF+ALKD
Sbjct: 421 LEDVSVKKLTRQNPTMREKFEALKD 445
```

SEQ ID 5846 (GBS304) was expressed in *E.coli* as a His-fusion product. The purified protein is shown in Figure 206, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1879

A DNA sequence (GBSx1987) was identified in *S.agalactiae* <SEQ ID 5849> which encodes the amino acid sequence <SEQ ID 5850>. This protein is predicted to be unnamed protein product. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1845(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5851> which encodes the amino acid sequence <SEQ ID 5852>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2492(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 113/180 (62%), Positives = 141/180 (77%)

Query: 16  LSELVDCFKGKAVPSKAEAGDIRIINLSDMSPLGIDYHNLRTFQDEQRSLLKYLLQEGDV 75
            L  +VDCFKGKAV SK   GD+ +INLSDM  LGI YH LRTFQ ++R LL+YLL++GDV
Sbjct: 18  LGTVVDCFKGKAVSSKVVPGDVGLINLSDMGTLGIQYHQLRTFQMDRRQLLRYLLEDGDV 77

Query: 76  LIASKGTVKKVAIFEEQDYPVVASANITILRPTQHIRGYYLKLFFDSEEGQQALENANKG 135
            LIASKGT+KKV +F +Q+  VVAS+NIT+LRP + +RGYY+K F DS  GQ  L+ A+ G
Sbjct: 78  LIASKGTLKKVCVFHKQNRDVVASSNITVLRPQKLLRGYYIKFFLDSPIGQALLDVADHG 137

Query: 136 KAVMNISTKELLNIAIPSIPLFRQDYLIQRYKQGLNDYKRKIARAEQEWERIQNDIRQQL 195
            K V+N+STKELL+I IP IPL +QDYLI  Y +GL DY RK+ RAEQEWE IQN+I++ L
Sbjct: 138 KDVINLSTKELLDIPIPVIPLVKQDYLINHYLRGLTDYHRKLNRAEQEWEYIQNEIQKGL 197
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1880

A DNA sequence (GBSx1988) was identified in *S.agalactiae* <SEQ ID 5853> which encodes the amino acid sequence <SEQ ID 5854>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -7.43   Transmembrane   62 -  78 (  55 -  82)
   INTEGRAL    Likelihood = -2.87   Transmembrane  130 - 146 ( 130 - 150)
   INTEGRAL    Likelihood = -1.28   Transmembrane   37 -  53 (  37 -  53)

----- Final Results -----
               bacterial membrane --- Certainty=0.3972(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9347> which encodes amino acid sequence <SEQ ID 9348> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA22372 GB:AL034446 putative transmembrane protein
          [Streptomyces coelicolor A3(2)]
 Identities = 38/139 (27%), Positives = 64/139 (45%), Gaps = 5/139 (3%)

Query: 15  SASVEILCRGWLLPVSATKYSKIVSVSISSIFFGLLHSANNHVSLISIFNLCL-FGLFLS 73
           +A+ E++ RG L  +        +++ ++ + FGL+H  N   +L     + +  G  L+
Sbjct: 143 AATEEVVFRGVLFRIIEEHIGTYLALGLTGLVFGLMHLLNEDATLWGALAIAIEAGFMLA 202

Query: 74  LYVILKGNIWGACGIHGAWNCVQGSVFGIEVSGEPMLSNSLVHVKTYGADWISGGKFGVE 133
                  N+W   G+H  WN   G VF   VSG    S  L+     G   ++GG FG E
Sbjct: 203 AAYAATRNLWLTIGVHFGWNFAAGGVFSTVVSGNGD-SEGLLDATMSGPKLLTGGDFGPE 261

Query: 134 GSMIT---SIVLIVACYWL 149
           GS+ +    ++L +   WL
Sbjct: 262 GSVYSVGFGVLLTLVFLWL 280
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1881

A DNA sequence (GBSx1989) was identified in *S.agalactiae* <SEQ ID 5855> which encodes the amino acid sequence <SEQ ID 5856>, which is a methylase gene homolog. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2192(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 264-266
```

A related GBS nucleic acid sequence <SEQ ID 9929> which encodes amino acid sequence <SEQ ID 9930> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA87672 GB:AB016260 Hypothetical gene, methylase gene homolog
          [Agrobacterium tumefaciens]
 Identities = 358/1238 (28%), Positives = 595/1238 (47%), Gaps = 99/1238 (7%)

Query: 1072 KEVARIKGMVDIRNAYQEVIAIQRYYDYDKETFNHLLGKLNRTYDSFVKHYGYLNSAV-- 1129
            K V  I+ ++ IR+A +EV+  Q         +  L  +L   + SFV+ +G +N
Sbjct: 497  KHVRIIRKLIPIRDAVREVLKAQEL----DRPWKDLQVRLRVAWSSFVRDFGPINHTTVS 552

Query: 1130 -----------------NRNLFDSDDKYSLLASLEDESL--DPSGKSVIYTKSLAFEKAL 1170
                             N   F  D    L+AS+ED  L  D +    I+T     E+ +
Sbjct: 553  ITEDPESGETRESHRRPNLQPFADDPDCWLVASIEDYDLENDTAKPGAIFT-----ERVI 607

Query: 1171 VRPEKEVKKVHTALDALNSSLADGRGVDFAYMMSIYQVESQMTLIEELGDLIMPDPEKYL 1230
              P   V  + +A DAL   L +   VD  ++  +   +    ++ ELG  I  DP
Sbjct: 608  SPPAPPV--ITSAADALAVVLNERGRVDLDHIAELLHRDPD-DVVAELGSAIFRDP---- 660
```

```
Query: 1231 NGELTYVSRQDFLSGDVVTKLEVVDLFVKQDNQDFNWSHYAGLLEAIKPARITLADIDYR 1290
              + ++     +LSG V  KL+V +     D     ++     L  ++P  +  +DI  R
Sbjct: 661  -ADGSWQMADAYLSGPVRDKLKVAEAAAALDPV---YNRNVTALAGVQPVDLRPSDITAR 716

Query: 1291 IGSRWIPLAVYGKFAQETFMGKAYELSDQ-EVATVLEVSPIDGVITYQSKFAYTYSNATD 1349
            +G+ WIP A    F +E  MG    +    E+A+    +   G +      A T    TD
Sbjct: 717  LGAPWIPAADVVAFVKE-MMGTDIRIHHMPELASWTVEARQLGYLA-----AGTSEWGTD 770

Query: 1350 RSLGVPASRYDSGRKIFENLLNSNQPTITKQVVEGDKKKNVTDVEKTTVLRAKETHLQEL 1409
            R             ++  + LNS  P I   + +GD ++ V +V  T   + K   +++
Sbjct: 771  RR---------HAGELLSDALNSRVPQIFDTIRDGDSERRVLNVVDTEAAKEKLHKIKDA 821

Query: 1410 FQGFVAKYPEVQQMIEDTYNRLYNRTVSKSYDGSHLTIDGLAQNISLRPHQKNAIQRIVE 1469
            FQ ++   P+    +   YN  +N    + + G HL + G +    L  HQK  I RI+
Sbjct: 822  FQRWIWSDPDRTDRLARVYNDRFNNIAPRKFSGDHLNLPGASGAFVLYGHQKRGIWRIIS 881

Query: 1470 EKRALLAHEVGSGKTLTMLGAGFKLKELGMVHKPLYVVPSSLTAQFGQEIMKFFPTKKVY 1529
                 LAH VG+GKT+TM  +  + + LG++ K + VVP    AQ  +E +  +PT ++
Sbjct: 882  SGSTYLAHAVGAGKTMTMAASIMEQRRLGLIAKAMQVVPGHCLAQAAREFLALYPTARIL 941

Query: 1530 VTTKKDFAKAKRKQFVSRIITGDYDAIVIGDSQFEKIPMSREKQVTYINDKLEQLREIKL 1589
            V  + +F+K KR +F+SR  T  +DAI+I  S F  I +    +   I+D+LE    + L
Sbjct: 942  VADETNFSKDKRARFLSRAATATWDAIIITHSAFRFIGVPAAFESQMIHDELELYETLLL 1001

Query: 1590 GSDSDYTV--KEAERSIKGLEHQLEELQKLERDTFIEFENLGIDFLFVDEAHHFKNIRPI 1647
              + +  V  K  ER  +GL+ +LE L    +D  +    +G+D + VDEA  F+ +
Sbjct: 1002 KVEDEDRVSRKRLERLKEGLQERLEALST-RKDDLLTIAEIGVDQIIVDEAQEFRKLSFA 1060

Query: 1648 TGLGNVAGITNTTSKKNVDMEMKVRQVQAEHGDRNVVFATGTPVSNSISELFTMMDYIQP 1707
            T +  + G+    S++  D+ +K R ++  +  R +V A+GTP++N++ E+F++   +
Sbjct: 1061 TNMSTLKGVDPNGSQRAWDLYVKSRFIETINPGRALVLASGTPITNTLGEMFSVQRLMGH 1120

Query: 1708 DVLERYLVSNFDSWVGAFGNIENSMELAPTGDKYQPKKRFKKFVNLPELMRIYKETADI- 1766
              LE    +  FD+W   FG+    +EL P+G KY+P  RF  FVN+PEL+ +++  AD+
Sbjct: 1121 AALEERGLHEFDAWASTFGDTTTELELQPSG-KYKPVSRFASFVNVPELIAMFRSFADVV 1179

Query: 1767 ---QTSDMLDLP-VPEAKIIAVESELTQAQKYYLEELVKRSDAIKSGS--VDPSRDNMLK 1820
                  + + +P +   +   V S+ TQA K++   L +R  AI+       P  D +L
Sbjct: 1180 MPADLREYVKVPAISTGRRQIVTSKPTQAFKHHQMVLAERIKAIEERERPPQPGDDILLS 1239

Query: 1821 ITGEARKLAIDMRLIDPTYSLSDNQKILQVVDNVERIYRDGAGDK-------------AT 1867
            +  + R  AID+RL+D       + K+  +V N  RI++  AG               A
Sbjct: 1240 VITDGRHAAIDLRLVDADNDNEPDNKLNNLVSNAFRIWKATAGSVYLRHDSKPFEVPGAA 1299

Query: 1868 QMIFSDIGTPK-SKEEGFDVYNELKDLFVDRGIPKEEIAFVHDANTDEKKNSLSRKVNSG 1926
            QMIFSD+GT    K  GF  Y  ++D  +  G+P  EIAF+ D    E K  L   V +G
Sbjct: 1300 QMIFSDLGTISVEKTRGFSAYRWIRDELIRLGVPASEIAFMQDFKKSEAKQRLFGDVRAG 1359

Query: 1927 EVRILMASTEKGGTGLNVQSRMKAVHYLDVPWRPSDIVQRNGRLIRQGNMHQEVDIYHYI 1986
             VR L+ S+E  GTG+NVQ R+KA+H+LDVPW PS I QR GR++RQGN H EVDI+ Y
Sbjct: 1360 RVRFLIGSSETMGTGVNVQLRLKALHHLDVPWLPSQIEQREGRIVRQGNQHDEVDIFAYA 1419

Query: 1987 TKGSFDNYLWQTQENKLKYITQIMTSKDPVRSAEDIDE-QTMTASDFKALATGNPYLKLK 2045
            T+GS D  +WQ  E K ++I   ++    +R  EDI E Q    +  KA+A+G+  L  K
Sbjct: 1420 TEGSLDATMWQNNERKARFIAAALSGDTSIRRLEDIGEGQANQFAMAKAIASGDQRLMQK 1479

Query: 2046 MELENELTVLENQKRAFNRSKDEYRHTISYSEKHLPIMEKRLSQYDKDIAQSLATKSQDF 2105
              LE ++  LE  + A    +   R  +  +E+ + +  +R+++  +DI + + T  +DF
Sbjct: 1480 AGLEADIARLERLRAAHIDDQHAVRRQLRDAERDIEVSTRRIAEIGQDITRLVPTTGEDF 1539

Query: 2106 VMRFDNQAMDNRAEAGDYLRK-LITYNRSETKEVRTLASFRGFDLKM-TTRGASEPLPET 2163
             M    +    R EAG  L K ++T  +   +    +AS  GF+L+    R   +    T
Sbjct: 1540 TMTVAGKDYSERKEAGRALMKEILTLVQLSPEGEAVIASIGGFELEYHGQRYGKDGYRYT 1599

Query: 2164 ISLMIVGDNQYTVALDLK-SDVGTIQRISNAIDHIIDDQEKTQELVKDLKDKLRVAKVEV 2222
              L   G + Y + L +  + +G + R+ +A+D    ++E+ ++ + D + +L   +
Sbjct: 1600 TMLKRTGAD-YEIELPVTVTPLGAVSRLEHALDDFDGERERYRQRLGDARRRLASYQSRG 1658
```

```
Query: 2223 DKVFPKEEDYQLVKAKYDVLAPLVEKEAEIEEIDAALA 2260
            +             +++     L EK  ++ E++ ALA
Sbjct: 1659 E------------GSEFAFAGELAEKHRQLAEVETALA 1684
 Identities = 99/271 (36%), Positives = 153/271 (55%), Gaps = 10/271 (3%)

Query: 607 RDKVETNIVAIRLVKNLEVEHRNASPSEQELLAKYVGWGG--LANEFFD-----DYNPKF 659
           +D+   NI AIRL   +E   R A+  EQE L ++ G+G   LAN  F      ++   +
Sbjct: 80  KDRARDNIAAIRLAAEIEASERPATREEQETLIRFTGFGASDLANGVFRRPGELEFRKGW 139

Query: 660 SKEREELKSLVTDKEYSDMKQSSLTAYYTDPSLIRQMWDKLERDGFTGGKILDPSMGTGN 719
            +   +L+  V + +Y+ + + +  A++T   ++R +W  L+R G+ GG++L+P +GTG
Sbjct: 140 DEIGSDLEDAVGETDYASLARCTQYAHFTPEFIVRAIWSGLQRLGWRGGRVLEPGIGTGL 199

Query: 720 FFAAMPKHLREKSELYGVELDTITGAIAKHLHPNSHIEIKGFETVAFNDNSFDLVISNVP 779
           F A MP+ LR+ S + GVELD +T  I + L P + I    F        SFDL I N P
Sbjct: 200 FPALMPEALRDLSHVTGVELDPVTACIVRLLQPRARILTGDFARTEL-PASFDLAIGNPP 258

Query: 780 FANIRIADNRYDRP--YMIHDYFVKKSLDLLHDGGQVAIISSTGTMDKRTENILQDIRET 837
           F++  +  +R  R     +HDYFV +S+DLL  G   A ++S+GTMDK      Q I  T
Sbjct: 259 FSDRTVRSDRAYRSLGLRLHDYFVARSIDLLKPGAFAAFVTSSGTMDKADSAARQHIATT 318

Query: 838 TEFLGGVRLPDSAFKAIAGTSVTTDMLFFQK 868
            + +  +RLP+ +F+A AGT V  D+LFF+K
Sbjct: 319 ADLIAAIRLPEGSFRADAGTDVVVDILFFRK 349
```

SEQ ID 5856 (GBS327N) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 148 (lane 8-10; MW 140kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 148 (lane 11-13; MW 115kDa) and in Figure 182 (lane 8; MW 115kDa).

Purified GBS327N-GST is shown in Figure 243, lane 5; Purified GBS327N-His is shown in Figure 235, lane 5.

GBS327C was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 148 (lane 14; MW 73kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1882

A DNA sequence (GBSx1990) was identified in *S.agalactiae* <SEQ ID 5857> which encodes the amino acid sequence <SEQ ID 5858>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3656(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1883

A repeated DNA sequence (GBSx1991) was identified in *S.agalactiae* <SEQ ID 5859> which encodes the amino acid sequence <SEQ ID 5860>. This protein is predicted to be giant membrane protein. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3698(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG19662 GB:AE005054 calcium-binding protein homology; Cbp
           [Halobacterium sp. NRC-1]
 Identities = 22/43 (51%), Positives = 29/43 (67%), Gaps = 1/43 (2%)

Query: 9   KDSDQDGLTDAQELAL-GTDPQSVDTDGDGQADLEELQSGHSP 50
           +D+D DGL+D  E+ + GTDP   DTDGDG  D  EL++G  P
Sbjct: 198 RDTDDDGLSDGVEVRVAGTDPTERDTDGDGVDDAAELRAGSLP 240
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1884

A DNA sequence (GBSx1992) was identified in *S.agalactiae* <SEQ ID 5861> which encodes the amino acid sequence <SEQ ID 5862>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.39    Transmembrane 1609 -1625 (1609 -1625)
   INTEGRAL    Likelihood = -1.81    Transmembrane   30 -  46 (  29 -  46)

----- Final Results -----
               bacterial membrane --- Certainty=0.1956(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif 1600-1604
```

The protein has homology with the following sequences in the GENPEPT database.

```
!GB:X57841 antigen I /II [Streptococcus sobrinus] (v...

>GP:CAA40973 GB:X57841 antigen I /II [Streptococcus sobrinus]
 Identities = 419/1436 (29%), Positives = 608/1436 (42%), Gaps = 310/1436 (21%)

Query: 23    KSKKYRTLCSVALGTMVTAVVAWGGTVAHADEVTTSV----DTTIQRTE--NPATNLPEA 76
              K K  RTL    LGT + A  A  G  A A+E +T+     DT +  TE  NPATNLP+
Sbjct: 23    KVKSGRTLSGALLGTAILASGA--GQKALAEETSTTSTSGGDTAVVGTETGNPATNLPDK 80

Query: 77    QPNP-------------------VSEQTESMASTGQSNGAIAVTVPHDTVT-----QAVE 112
              Q NP                   V   T  +    +S     VTV  D        +  +
Sbjct: 81    QDNPSSQAETSQAQARQKTGAMSVDVSTSELDEAAKSPQEAGVTVSQDATVNKGTVEPSD 140

Query: 113   EAKAEGVSTVEDSPMDLGNTRSAVET---------------NQQIS-------------K 144
              EA  +     +D      + + A E                NQ+I+             K
Sbjct: 141   EANQKEPEIKDDYSKQAADIQKATEDYKASVAANQAETDRINQEIAAKKAQYEQDLAANK 200
```

```
Query: 145  AD-------------------ADTQKQVETINEVTK----TYKADKATYESNKARIEQEN 181
            A+                   A  QK +  I +        Y A K  Y+   AR++  N
Sbjct: 201  AEVERSLMRMRKPRPIYEAKLAQNQKDLAAIQQANSDSQAAYAAAKEAYDKEWARVQAAN 260

Query: 182  KELSQAYEGANQTGKETNAWVDTKVNDLKARYADADVTVKEQ-------VVSSGNGTSVL 234
                +AYE A       N  +  ++  ++ R A AD   K            +GN  +
Sbjct: 261  AAAKKAYEEALAANTAKNDQIKAEIEAIQQRSAKADYEAKLAQYEKDLAAAQAGNAANEA 320

Query: 235  DY----TNYGKAVETIQSTNEQAVADY----LTKKTKADDIVAKNQAIQKENEA------ 280
            DY      Y + +  +Q+ N  A   Y         K   I A+N+AIQ+  +A
Sbjct: 321  DYQAKKAAYEQELARVQAANAAAKQAYEQALAANSAKNAQITAENEAIQQNAQAKADYEA 380

Query: 281  -------GLANAKADNEAIERRNQAGQAAVDAEN---RAGQAAVDQANQEKQQLVSDRAA 330
                    LA A++ N A E   Q   AA + E    +A  AA  QA +++ Q  + + A
Sbjct: 381  KLAQYQKDLAAAQSGNAANEADYQEKLAAYEKELARVQAANAAAKQAYEQQVQQANAKNA 440

Query: 331  EIEAITKRNKEKEAAARKENEAIDAYNTKEMERYQRDLAEIS------------------ 372
            EI    +  +E+ A A+ + E   +   +E+ +Y++DLAE
Sbjct: 441  EITEANRAIRERNAKAKTDYELKLSKYQEELAQYKKDLAEYPAKLQAYQDEQAAIKAALA 500

Query: 373  -----KGEEGYISEALAQALNLNNGEPQAQHGAITRN----------------------- 404
                 K E+G +SE  AQ+L + + EP AQ   +T
Sbjct: 501  ELEKHKNEDGNLSEPSAQSL-VYDLEPNAQVALVTDGKLLKASALDEAFSHDEKNYNNHL 559

Query: 405  --PDQI----------ISTGDALLGGYSRILDSTGF-----------FVYDMFKTGETLS 441
              PD +          +++   L G +    D  G+           F   + K G++ +
Sbjct: 560  LQPDNLNVTYLEQADDVASSVELFGNFG---DKAGWTTTVSNGAEVKFASVLLKRGQSAT 616

Query: 442  FNYQNLQHARFDGKKISRVTYDITNLVSPAG-----TNAVKLVVPNDPTEGFIAYRNDGN 496
              Y NL+++ ++GKKIS+V Y  T  V P       T  V L +  DPT G  A    G
Sbjct: 617  ATYTNLKNSYYNGKKISKVVYKYT--VDPDSKFQNPTGNVWLGIFTDPTLGVFASAYTGQ 674

Query: 497  GDWRTD---KMEFRVVAKYYLEDGSQVTFSKEKPGVFTHSSLNHNDIGLEYVKDSSGKFV 553
             +  T    K EF     +Y EDG+ + F      + + +SLN     +E  KD SG FV
Sbjct: 675  NEKDTSIFIKNEF----TFYDEDGNPIDFDN---ALLSVASLNREHNSIEMAKDYSGTFV 727

Query: 554  PINGSTVQVTN--------------EGLARSLGSNRASDLNLPEEWDTTSSRYAYKGAIV 599
             I+GS++   N              EG +      RAS+      WD+  +  ++ GA
Sbjct: 728  KISGSSIGEKNGMIYATDTLNFKKGEGGSLHTMYTRASEPG--SGWDSADAPNSWYGAGA 785

Query: 600  STVTSGNTY--------TVTFGQGDMPQNVGL--------SYWFALN------------- 630
              ++  N Y        T      +MPQ  G         + W++LN
Sbjct: 786  VRMSGPNNYITLGATSATNVLSLAEMPQVPGKDNTAGKKPNIWYSLNGKIRAVNVPKVTK 845

Query: 631  --TLPVARTVTPYSPKPHVTVEL-----EPIPEPITVTPDIYTPKTFTPEKPVTFT---- 679
                P      P  P   V  EL     EP  EP   TP    P    PEKPV  T
Sbjct: 846  EKPTPPVEPTKPDEPTYEVEKELVDLPVEPKYEP-EPTPPSKNPDQSIPEKPVEPTYEVE 904

Query: 680  ----PKPLDEVVQPSLTLTKVT-------LPVKPIPKELPTPP------------QVPTV 716
                P P++   +   T  + T        PV+P  + LPTPP             VPTV
Sbjct: 905  KELEPAPVEPSYEKEPTPPQSTPDQEEPEKPVEPSYQSLPTPPVEPVYETVPGPVSVPTV 964

Query: 717  HYHAYRLTTTSEIMKEVVNSDQANLHEKTVAKDSTVIYPLTVDALSPNRAQTTSLIFEDY 776
             YH Y+L     + KE+ N D  ++ +  VAK STV + L    L   R +TTS +  D
Sbjct: 965  RYHYYKLAVQPGVTKEIKNQDDLDIDKTLVAKQSTVKFQLKTADLPAGRPETTSFVLMDP 1024

Query: 777  LPAGYLFDKETTQKENGNYVLSFDETKNFVTLTAKENLLQEVNKDLTQVYQLTAPKLYGS 836
            LP+GY  + E T+  +  +  S+D   + VT TA    L  +N+DLT+      P + G
Sbjct: 1025 LPSGYQLNLEATKVASPGFEASYDAMTHTVTFTATAETLAALNQDLTKAVATIYPTVVGQ 1084

Query: 837  VQNDGATYSNSYKLLLNKGTTNAYTVTSNVVTVRTPG-----DGETTTLITPDKNNENAD 891
            V NDGATY+N++ L++N    +AY + SN+V V TPG     D  +   ITP K N+N +
Sbjct: 1085 VLNDGATYTNNFTLMVN----DAYGIKSNIVRVTTPGKPNDPDNPSNNYITPHKVNKNEN 1140

Query: 892  GVLINDTVVALGTTNHYRLTWDLDQYKGDRSAKETIARGFFFVDDYPEEVLDVVENGTAI 951
            GV+I+   V  GTTN+Y LTWDLDQYKGD+SAKE I +GFF+VDDYPEE LD+  +   +
Sbjct: 1141 GVVIDGKSVLAGTTNYYELTWDLDQYKGDKSAKEIIQKGFFYVDDYPEEALDLRTDLIKL 1200
```

```
Query: 952  TTLDGQKVSGITVKNYASLNEAPKDLQDKLARAKITPTGAFQVFMPDDNQAFYDQYVQTG 1011
            T  +G+ V+G++V +YASL  AP  +QD L +A I P GAFQVF  DD QAFYD YV TG
Sbjct: 1201 TDANGKAVTGVSVADYASLEAAPAAVQDMLKKANIIPKGAFQVFTADDPQAFYDAYVVTG 1260

Query: 1012 TSLALLTKMTVKDSLYGQTKTYTNKAYQVDFGNGYETKEVTNTLVSPEPKKQ-NLNKDKV 1070
            T L ++T MTVK  +     +Y N+AYQ+DFGNGYE+  V N +    P+K   L  D
Sbjct: 1261 TDLTIVTPMTVKAEMGKTGGSYENRAYQIDFGNGYESNLVVNNVPKINPEKDVTLTMDPA 1320

Query: 1071 D---INGKPMLVGTQNHYTLSWDLDQYRGIKADNSQIAQGFYFVDDYPE-----EALLPD 1122
            D   ++G+ + +    +Y L   +     I AD+++    + F DDY +
Sbjct: 1321 DSTNVDGQTIALNQVFNYRLIGGI-----IPADHAEELFEYSFSDDYDQTGDQYTGQYKA 1375

Query: 1123 EAAIQFVTSDGKTV-SGITVKSY--SQLLEAPKTLQAAFSKQKIQPKGAFQVFMPE 1175
             A +     DG  + +G  + SY  +Q+ EA   +   F +  ++       F  E
Sbjct: 1376 FAKVDLTLKDGTIIKAGTDLTSYTEAQVDEANGQIVVTFKEDFLRSVSVDSAFQAE 1431
 Identities = 209/442 (47%), Positives = 280/442 (63%), Gaps = 27/442 (6%)

Query: 1198 TVLETMLNSGKSY-ENVAYQVDFGQAYETNTVTNFVPK------------VTPHKSNTNQ 1244
            TV+  +LN G +Y  N    V+     ++N V    P             +TPHK N N+
Sbjct: 1080 TVVGQVLNDGATYTNNFTLMVNDAYGIKSNIVRVTTPGKPNDPDNPSNNYITPHKVNKNE 1139

Query: 1245 EGISIDGKTVLPNTVNYYKIVLDYSQYKDMVVTDDVLAKGFYMVDDYPEEALTLNPDGIQ 1304
             G+ IDGK+VL  T NYY++  D  QYK      +++ KGF+ VDDYPEEAL L  D I+
Sbjct: 1140 NGVVIDGKSVLAGTTNYYELTWDLDQYKGDKSAKEIIQKGFFYVDDYPEEALDLRTDLIK 1199

Query: 1305 VLDKDGNRVSGISVSTYASLSEAPKVVQDAMAKRQFTPKGAIQVLSSDDPKVFYDTYVKT 1364
            + D +G  V+G+SV+ YASL  AP  VQD + K    PKGA QV ++DDP+ FYD YV T
Sbjct: 1200 LTDANGKAVTGVSVADYASLEAAPAAVQDMLKKANIIPKGAFQVFTADDPQAFYDAYVVT 1259

Query: 1365 GQTLVVTLPMTVKNELTKTGGQYENTAYQIDFGLAYVTETVVNNVPKLDPQKDVVIDLSH 1424
            G  L +  PMTVK E+ KTGG YEN AYQIDFG  Y +  VVNNVPK++P+KDV + +
Sbjct: 1260 GTDLTIVTPMTVKAEMGKTGGSYENRAYQIDFGNGYESNLVVNNVPKINPEKDVTLTMDP 1319

Query: 1425 KDA-SLDGKEVALHQTFNYRLVGAMIPSNRATDLFEYGFEDNYDEKHDEYNGVYRSYLMT 1483
             D+ ++DG+ +AL+Q FNYRL+G +IP++ A +LFEY F D+YD+  D+Y G Y+++
Sbjct: 1320 ADSTNVDGQTIALNQVFNYRLIGGIIPADHAEELFEYSFSDDYDQTGDQYTGQYKAFAKV 1379

Query: 1484 DVILKDGSVLKEGTEVTKYTLQQVDTENGLVSISFDKSFLETVSDDSAFQADVYLQMKRI 1543
            D+ LKDG+++K GT++T YT  QVD  NG + ++F + FL +VS DSAFQA+VYLQMKRI
Sbjct: 1380 DLTLKDGTIIKAGTDLTSYTEAQVDEANGQIVVTFKEDFLRSVSVDSAFQAEVYLQMKRI 1439

Query: 1544 AAGQVENTYLHTVNGYVISSNTVVTHTPQPEEPSPNQP--------TPPQPPIETIEPPV 1595
            A G   NTY++TVNG   SSNTV T TP+P++PSP  P         P Q       PP
Sbjct: 1440 AVGTFANTYVNTVNGITYSSNTVRTSTPEPKQPSPVDPKTTTTVVFQPRQGKAYQPAPPA 1499

Query: 1596 PASILPNTGEQES----LLGLI 1613
             A  LP TG+  +    LLGL+
Sbjct: 1500 GAQ-LPATGDSSNAYLPLLGLV 1520
 Identities = 100/210 (47%), Positives = 137/210 (64%), Gaps = 4/210 (1%)

Query: 1060 PKKQNLNKDKVDINGKPMLVGTQNHYTLSWDLDQYRGIKADNSQIAQGFYFVDDYPEEAL 1119
            P K N N++ V I+GK +L GT N+Y L+WDLDQY+G K+    I +GF++VDDYPEEAL
Sbjct: 1132 PHKVNKNENGVVIDGKSVLAGTTNYYELTWDLDQYKGDKSAKEIIQKGFFYVDDYPEEAL 1191

Query: 1120 LPDEAAIQFVTSDGKTVSGITVKSYSQLLEAPKTLQAAFSKQKIQPKGAFQVFMPEDPQA 1179
                  I+   ++GK V+G++V  Y+ L  AP  +Q    K  I PKGAFQVF  +DPQA
Sbjct: 1192 DLRTDLIKLTDANGKAVTGVSVADYASLEAAPAAVQDMLKKANIIPKGAFQVFTADDPQA 1251

Query: 1180 FFESYVTKGENITIVTPMTVLETMLNSGKSYENVAYQVDFGQAYETNTVTNFVPKVTPHK 1239
            F+++YV  G ++TIVTPMTV   M  +G SYEN AYQ+DFG  YE+N V N VPK+ P K
Sbjct: 1252 FYDAYVVTGTDLTIVTPMTVKAEMGKTGGSYENRAYQIDFGNGYESNLVVNNVPKINPEK 1311

Query: 1240 SNT----NQEGISIDGKTVLPNTVNYYKIV 1265
              T      +  ++DG+T+  N V  Y+++
Sbjct: 1312 DVTLTMDPADSTNVDGQTIALNQVFNYRLI 1341
```

There is also homology to SEQ ID 598.

SEQ ID 5862 (GBS76) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 2; MW 17.4kDa). The GBS76-His fusion product was purified (Figure 196, lane 8) and used to immunise mice. The resulting antiserum was used for FACS (Figure 294), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1885

A DNA sequence (GBSx1993) was identified in *S.agalactiae* <SEQ ID 5863> which encodes the amino acid sequence <SEQ ID 5864>. This protein is predicted to be abortive infection bacteriophage resistance protein (abiEi). Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2765(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9931> which encodes amino acid sequence <SEQ ID 9932> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB52382 GB:U36837 AbiEi [Lactococcus lactis]
 Identities = 51/206 (24%), Positives = 90/206 (42%), Gaps = 23/206 (11%)

Query: 17  KNNGIVTNKDCKALGIPTIYLTRLEKEGIIFRVEKGIFLTQNGDYDEYYFFQYRFPKAIF 76
           K  G +  K  +  GI   YL +   +  +  V+KG+++  +   D  + FQ ++ KA+
Sbjct: 76  KYKGNIIRKIVRDEGISDYYLRKFVLKYNLTEVDKGVYIFPHKKKDSLFIFQQKYSKAVI 135

Query: 77  SYISALYLQQFTDEIPQYFDVTVPRGYRF----------------------NTPPANLNI 114
           S+ ++LYLQ   D IPQ   ++VP  Y                        N    N+ I
Sbjct: 136 SHETSLYLQDVIDYIPQKIQMSVPEKYNISRIQEPHENRLTSYNYVDINSNNIMDKNIPI 195

Query: 115 HFV-SKEYSELGMTTVPTPMGNNVRVYDFERIICDFVIHREKIDSELFVKTLQSYGNYPK 173
           + V +K  S   + TV + +G  +RV    R I D +    K + E+  + ++ Y
Sbjct: 196 NLVRNKSISPTQIETVNSFLGLPLRVTSIARSIVDVLKPSHKAEEEVKEQAIKYYLERFP 255

Query: 174 KNLAKLYEYATKMNTLEKVKQTLEVL 199
            N+ +L   A   N L++++  L +L
Sbjct: 256 DNIVRLKRIAKTQNVLKELEYYLILL 281
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1886

A DNA sequence (GBSx1994) was identified in *S.agalactiae* <SEQ ID 5865> which encodes the amino acid sequence <SEQ ID 5866>. This protein is predicted to be abortive infection bacteriophage resistance protein (abiEii). Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.12   Transmembrane  260 - 276 ( 259 - 277)
```

```
----- Final Results -----
               bacterial membrane --- Certainty=0.1447(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB52383 GB:U36837 AbiEii [Lactococcus lactis]
 Identities = 76/276 (27%), Positives = 135/276 (48%), Gaps = 19/276 (6%)

Query: 14  SKNTGLTFNSVMTYYFLEVILKKLSQSSYSNHYIFKGGFLLSNVIGVESRSTVDIDFLFH 73
           ++N  +   +    Y  E  L +LS S Y   ++ KGGFL+     +  R+T D+D
Sbjct: 12  TRNDDIGIENYRIRYATERFLTRLSASQYKEKFVLKGGFLIGVTYNLSQRTTKDLDTALI 71

Query: 74  QITLSEETVKQQLKEIL-ADSEEGISFVIQSITTIKESDDYGGYRATISCQLE--NIKQV 130
                 +++++ + EI   D E+ + F ++ +T+ ++   Y GYRA +       N +
Sbjct: 72  DFKSDAQSIERVITEICNIDLEDQVLFKLKELTSSQDMRIYPGYRAKLKMMFPDGNTRID 131

Query: 131 IHLDIATGDVVTPQPITYDYKAIFDE-----DNFPIIAYTIETILAEKLQTIYSRNFLNS 185
             LDI  GD +TP+         IF+E          ++AY  ETI AEKL+TI +R  +N+
Sbjct: 132 FDLDIGVGDRITPEAKKIKIPLIFNEVKGVEKQIEVLAYPKETIQAEKLETILTRGKVNT 191

Query: 186 RSKDFYDVYIL--SKLKKKDIDFNQLKNACQRTFSYRE-TELDFEKIIE-----LLERFK 237
           R KD+YD ++L   +     I F     A + T+ +R  T+   E++ E     L E  +
Sbjct: 192 RMKDYYDFHLLLTDQENSNSISFYY---AFKNTWEFRNPTQFIDEELFEDWLFILDEILE 248

Query: 238 SDPTQNQQWQNYSKKYSYTKGISLANVLDEMISLIT 273
           S   + + W NY K  +Y K +++ +++ E+   ++
Sbjct: 249 SKELKEKYWPNYIKDRNYAKHLNMDDIISEIKEFVS 284
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1887

A DNA sequence (GBSx1995) was identified in *S.agalactiae* <SEQ ID 5867> which encodes the amino acid sequence <SEQ ID 5868>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1137(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1888

A DNA sequence (GBSx1996) was identified in *S.agalactiae* <SEQ ID 5869> which encodes the amino acid sequence <SEQ ID 5870>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2782(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1889

A DNA sequence (GBSx1997) was identified in *S.agalactiae* <SEQ ID 5871> which encodes the amino acid sequence <SEQ ID 5872>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.14   Transmembrane  310 - 326 ( 301 - 334)

----- Final Results -----
               bacterial membrane --- Certainty=0.5055(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG38044 GB:AF295925 Orf28 [Streptococcus pneumoniae]
 Identities = 272/344 (79%), Positives = 307/344 (89%)

Query: 568 VYVNPAFYFPKVIQVQTTILPTIGQFGGDEFERAKAIYDYLKSKGATNQAIAAILGNWSV 627
            +YVNP FYFPKVIQ+QTTILP IGQFGGDEFERAK IY++LKS+GA+ QAIAAILGNWSV
Sbjct: 1   MYVNPQFYFPKVIQLQTTILPAIGQFGGDEFERAKHIYEFLKSQGASPQAIAAILGNWSV 60

Query: 628 ESSINPKRAEGDYLSPPVGATDSSWDDEGWLTLNGPTIYNGRYPNILKRGLGLGQWTDTA 687
            ESSINPKRAEGDYL+PPVG     WDDE WL + GP IY+G YPNIL RGLGLGQWTDTA
Sbjct: 61  ESSINPKRAEGDYLTPPVGVPIPPWDDESWLAIGGPAIYSGAYPNILHRGLGLGQWTDTA 120

Query: 688 DGSRRHTLLLEYAKGKHQKWYDLGLQLDFMLYGDSPYYTNWLKDFFKNSGSPASLAQLFL 747
            DGS RHT LL YA+ +++KWYDL LQLDFML+GDSPYY +WLKDFFKN+GS A+LAQLFL
Sbjct: 121 DGSTRHTALLNYARTQNKKWYDLDLQLDFMLHGDSPYYQSWLKDFFKNTGSAANLAQLFL 180

Query: 748 IYWEGNSGDKLLERQTRASEWYYQIEKGFSQPNGGTAQSDPKALEAVREDLFENSIPGGG 807
             YWEGNSGDKLLERQTRA+EWYYQIEKGFSQ NGG A+SDP++LE VR DL+++S+PGGG
Sbjct: 181 TYWEGNSGDKLLERQTRATEWYYQIEKGFSQTNGGQAKSDPQSLEGVRGDLYDHSVPGGG 240

Query: 808 DGMGYAYGQCTWGVAARINQLGLKLKGKNGEKIPIISTMGNGQDWVRTAASLGGETGTSP 867
            DGM YAYGQCTWGVAAR+NQLGLKLKG+NGEKI II+TMGNGQDWV T++SLGGETG++P
Sbjct: 241 DGMAYAYGQCTWGVAARMNQLGLKLKGRNGEKISIINTMGNGQDWVATSSSLGGETGSTP 300

Query: 868 QEGAILSFAGGGHGTPTEYGHVAFVEKVYPDGSFLISETNYNGN 911
            + GAI+SF GG HGTP  YGHVAFVEKVY DGSFL+SETNY GN
Sbjct: 301 RAGAIVSFVGGTHGTPASYGHVAFVEKVYDDGSFLVSETNYGGN 344
```

SEQ ID 5872 (GBS74d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 121 (lane 3 & 4; MW 95.5kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 121 (lane 5-7; MW 70.5kDa) and in Figure 179 (lane 9; MW 70.5kDa).

GBS74d-His was purified as shown in Figure 233, lane 7-8.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1890

A DNA sequence (GBSx1998) was identified in *S.agalactiae* <SEQ ID 5873> which encodes the amino acid sequence <SEQ ID 5874>. This protein is predicted to be TrsE-like protein. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5526(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG38042 GB:AF295925 Orf26 [Streptococcus pneumoniae]
 Identities = 618/782 (79%), Positives = 712/782 (91%), Gaps = 1/782 (0%)

Query: 1   MKKLKHSMKSK-TSSNDKKQKTKTQKQEISPSTVNTLAYQGLFQNGLMQVSPSYFSQTYL 59
           MK+  +++K + TS+ +KK++ K +K+E+ PST NTL+YQ L+QNGLMQV   YFSQ+YL
Sbjct: 3   MKRKSNTLKKQQTSTTNKKEEVKDKKEEVLPSTANTLSYQALYQNGLMQVKEDYFSQSYL 62

Query: 60  LGDVNYQTVGLDDKGAIVEKYSDLINSLDDKTNFQLTIFNQKVNLEKFRKSILYPLQEDG 119
           LGDVNYQTVGL+DKGAI+EKYSDLI SLDD+TNFQLTIFN+++NLEKFR S+LY  +EDG
Sbjct: 63  LGDVNYQTVGLEDKGAIIEKYSDLIKSLDDQTNFQLTIFNKRLNLEKFRHSVLYEEKEDG 122

Query: 120 FDTYRDELNRMMDANLEAGENNFSAVKFLSFGKSDQTPKLAFRSLSQIGEYFKSGFSEID 179
           +D+YR ELNRMM+ NL++GENNFSAVK +SFG+ D  PK A+RSLSQIGEYFKSGFSEID
Sbjct: 123 YDSYRKELNRMMNQNLDSGENNFSAVKLISFGRKDSNPKQAYRSLSQIGEYFKSGFSEID 182

Query: 180 VSLGLLGGEERVNVLADMLRGENHLPFSYKDLTLSGQSTKHFIAPTYLSFKHKNHIELDD 239
                L GEERVN+LADMLRGE+HLPFSY+DLT SGQ+T+HFIAP  L FK+KN+++++D
Sbjct: 183 ARFESLAGEERVNLLADMLRGEHHLPFSYRDLTRSGQTTRHFIAPNLLDFKNKNYLQIND 242

Query: 240 RLLQIVYVRDYGMELGDKFIRDLMQSDLEVMISLHAKGSTKSETMTKLRTKKTLMESQKI 299
           RLLQIVYVRDYGMELGD+FIRDLMQ DLE+++SLHA+ STKS+ M KLRTKKTLMESQKI
Sbjct: 243 RLLQIVYVRDYGMELGDQFIRDLMQGDLELIVSLHAQSSTKSDAMKKLRTKKTLMESQKI 302

Query: 300 GEQQKMARTGIYLEKVGHVLENNIDEAEALLQTMTQTGDKLFDTVFLIGVLADTEDQLKQ 359
           GEQQK+ARTGIYLEKVGHVLE+NIDEAE LL+TMT+TGDKLF TVFLIGV    E++LKQ
Sbjct: 303 GEQQKLARTGIYLEKVGHVLESNIDEAEELLKTMTETGDKLFQTVFLIGVFGQDEEELKQ 362

Query: 360 SLDIIKQVAGSNDMIIDNLTYMQEAAFNSLLPFGKNYLEGVSRSLLTSNIAVNAPWTSVD 419
           +LD ++QVAGSND++ID L YMQEAAFNSLLPFG ++LEGVSRSLLTSNIAVN+PWTSVD
Sbjct: 363 ALDTVQQVAGSNDLMIDKLPYMQEAAFNSLLPFGCDFLEGVSRSLLTSNIAVNSPWTSVD 422

Query: 420 IHDKGGKFYGINQISSNIISIDRGKLNTPSGLILGTSGAGKGMATKHEIISTKLKEADSD 479
           + D+ GK+YGINQISSNII+IDR  LNTPSGLILGTSGAGKGMATKHEII+TK+KE+  +
Sbjct: 423 LQDRSGKYYGINQISSNIITIDRSLLNTPSGLILGTSGAGKGMATKHEIITTKIKESGEN 482

Query: 480 TEIIIVDPENEYSIIGQAFGGESIDIAPDSTTFLNVLELSDENMDEDPVKVKSEFLLSWI 539
           TEIIIVDPE EYS+IG+ FGGE IDIAPDS T+LNVL+LS+ENMDEDPVKVKSEFLLS+I
Sbjct: 483 TEIIIVDPEAEYSVIGRTFGGEMIDIAPDSETYLNVLDLSEENMDEDPVKVKSEFLLSFI 542

Query: 540 GKLLDRKMDGREKSLIDRVTRLTYKHFDTPSLVEWVFVLSQQPEQEAKDLALDMELYVEG 599
           GKLLDRKMDGREKS+IDRVTRLTY+ F  PSL EWVFVLSQQPE+EA++LALDMELYVEG
Sbjct: 543 GKLLDRKMDGREKSIIDRVTRLTYQSFKEPSLEEWVFVLSQQPEEEAQNLALDMELYVEG 602

Query: 600 SLDIFSHRTNIKTDSHFLIYNVKKLGDELKQIALMVIFDQIWNRVVKNQKLGKKTWIYFD 659
           SLDIFSH+TNI+T S+FLIYNVKKLGDELKQIALMV+FDQIWNRVV+NQKLGKKTWIYFD
Sbjct: 603 SLDIFSHKTNIQTGSNFLIYNVKKLGDELKQIALMVVFDQIWNRVVRNQKLGKKTWIYFD 662
```

```
Query: 660 EMQLLLLDKYASDFFFKLWSRVRKYGAIPTGITQNVETLLLDANGRRIIANSEFMILLKQ 719
           E++LLLLDKY SDFFFKLWSRVRKYGA PTGITQNVETLLLD NGRRIIANSEFMILLKQ
Sbjct: 663 EIELLLLDKYPSDFFFKLWSRVRKYGASPTGITQNVETLLLDPNGRRIIANSEFMILLKQ 722

Query: 720 AKSDREELVHMLGLSKELEKYLVNPEKGAGLIKAGSTVVPFKNKIPQHTKLFDIMSTDPE 779
           AK+DREELV +LGLSKELEKYLVNPEKGAGLIKAGS VVPFKNKIPQ ++LFDIM +DP+
Sbjct: 723 AKNDREELVQLLGLSKELEKYLVNPEKGAGLIKAGSVVVPFKNKIPQGSQLFDIMRSDPD 782

Query: 780 KM 781
           KM
Sbjct: 783 KM 784
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8925> and protein <SEQ ID 8926> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:    -26.26
GvH: Signal Score (-7.5): -3.87
     Possible site: 55
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 0 value:   6.26 threshold:  0.0
   PERIPHERAL  Likelihood =  6.26     335
 modified ALOM score:  -1.75

*** Reasoning Step: 3

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.5526(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear)
```

The protein has homology with the following sequences in the databases:

```
33.5/57.2% over 789aa
                                                 Enterococcus faecalis
  GP|8100663| TrsE-like protein Insert characterized

ORF01332(319 - 2628 of 2949)
GP|8100663|gb|AAF72347.1|AF192329_8|AF192329(2  -  791  of  799)  TrsE-like  protein
{Enterococcus faecalis
}
%Match = 20.7
%Identity = 33.4  %Similarity = 57.2
Matches = 259  Mismatches = 323  Conservative Sub.s = 184

210       240       270       300       330       360                 387
SCYLGSIAPTIYHLKYTSSTVFIMN*RCQTAHLLEEKETNVKKLKHSMKSKTSSNDKKQKTKTQKQEI-----------S
                                              | |      :| : |: |::||
                                             MSKKEIPRETEKTKLTRAQRKEIDAVIRKYKGDGR
                                                     10        20        30

414       444       474       504       534       564       594       624
PSTVN-TLAYQGLFQNGLMQVSPSYFSQTYLLGDVNYQTVGLDDKGAIVEKYSDLINSLDDKTNFQLTIFNQKVNLEKFR
| |   :: |: :: :|: :|||  ||:   : |::||    | : || ||  || | :|   : |:: :|:||:  ::
PHTAQQSIPYEVMYPDGVCRVSPGVFSKCIEFADISYQLAQPDTQTAIFEKLCDLYNYVDASIHIQFSFLNRKVDPVQYA
             50        60        70        80        90       100       110

654       684       714       744       774       804       834       864
KSILYPLQEDGFDTYRDELNRMMDANLEAGENNFSAVKFLSFGKSDQTPKLAFRSLSQIGEYFKSGFSEIDVSLGLLGGE
||     | | ||  | |   ::   |  | |     |:|:|    :: | |   | :||  :   |  :     :: |
KSFEIAPQGDDFDDIRAEYTGILQKQLANGNNGMVKTKYLTFTIEAESVKAARARLKRIGFDLLGYFKSMGAVAHVMDGW
```

```
            130       140       150       160       170       180       190

894       921       951       981       1011      1041      1071      1101
ERVNVLADMLRGENHL-PFSYKDLTLSGQSTKHFIAPTYLSFKHKNHIELDDRLLQIVYVRDYGMELGDKFIRDLMQSDL
||:|:|  :   :  :  | :| |  || ||| ||||: | | :     :  :   : :::    || |  : |:: ::
ERLNLLHGVYHPDGEIFNFDWKWLAPSGLSTKDFIAPSSLCFGNAKTFGMGGKYGAVSFLQILSPELSDDMLADFLNTES
            210       220       230       240       250       260       270

1131      1161      1191      1221      1251      1281      1311      1341
EVMISLHAKGSTKSETMTKLRTKKTLMESQKIGEQQKMARTGIYLEKVGHVLENNIDEAEALLQTMTQTGDKLFDTVFLI
 |:::|| :   ::: : :: | | ::: || ||:|  |:|  ::  :  |    ::|: ||  :    ::||   ||:
GVLVNLHVQAIEQTKAIKTIKRKITDLDAMKIAEQKKAVRSGYDMDILPSDLATYGEDAKKLLTKLQTRNERLFQLTFLV
            290       300       310       320       330       340       350

1371      1401      1431      1461      1491      1521      1551
GVLADTEDQLKQSLDIIKQVAGSNDMIIDNLTYMQEAAFNSLLPFGKNYLEGVSRSLLTSNIAVNAPWTSVDIHDKGGK-
  :|||: :|   :     ||  ::  :  | | ||  : | ||:| | :  : ||| ||::||  |: : ::   |
LNVADTKQKLNNDVFQAAGVAQKHNCPLVRLDYQQEQGLASSLPLGVNQI-KIQRSLTTSSVAVFVPFVTQELFQGGAAM
            370       380       390       400        410       420       430

1608      1638      1668      1698      1728      1758      1788      1818
FYGINQISSNIISIDRGKLNTPSGLILGTSGAGKGMATKHEIISTKLKEADSDTEIIIVDPENEYSIIGQAFGGESIDIA
:||||  | |:| :|| :   |: | ||| |:|| |: | ||:|  |   |   :| | ||| ||  : :  : |: | ::
YYGINAKSRNMIMLDRKQARCPNALKLGTPGSGKSMSCKSEIVSVFLTTPD---DIFISDPEAEYYPLVKRLHGQVIRLS
             450       460       470       480          490       500       510

1848      1875      1905      1935      1959      1989      2019
PDSTTFLNVLELS-DENMDEDPVKVKSEFLLSWIGKLLDRK--MDGREKSLIDRVTRLTYKHFDTPSLVEWVFVLS----
| |  |:| |::: : : |::|: :||:|:||:   ::   |  ::  ||::||  |: | : :      |  : :||
PTSKDFVNPLDINLNYSEDDNPLALKSDFVLSFCELVMGGKNGLEAIEKTVIDRAVRVIYRPYLADPRPENMPILSDLHK
                530       540       550       560       570       580       590

2058      2088      2118      2148      2178      2208      2238      2268
---QQPEQEAKDLALDMELYVEGSLDIFSHRTNIKTDSHFLIYNVKKLGDELKQIALMVIFDQIWNRVVKNQKLGKKTWI
    |   ||  :|  ::||| |||::|:||||:   : :: :::|:|| :||:: :::: |||| ||  |:  || ||
ALLDQHVPEADRVAQALDLYVSGSLNVFNHRTNVDIGNRLVSFDIKELGKQLKKLGMLIVQDQIWGRVTANRSQGKATWY
                610       620       630       640       650       660       670

2298      2328      2358      2388      2418      2448      2478      2508
YFDEMQLLLLDKYASDFFFKLWSRVRKYGAIPTGITQNVETLLLDANGRRIIANSEFMILLKQAKSDREELVHMLGLSKE
: || :||| ::  : :  ::| | ||:| |||| ||||: ||       |: ||:|:| || ||  ||: |   | || |
FADEFHLLLKEEQTAAYSAEIWKRFRKWGGIPTGATQNVKDLLSSPEIENILENSDFITLLNQASGDRKILAERLNLSTE
                690       700       710       720       730       740       750

2538      2568      2598      2628      2658      2688      2718      2748
LEKYLVNPEKGAGLIKAGSTVVPFKNKIPQHTKLFDIMSTDPEKMRT*DERG*KASQTG*AKLSKQLKISSYALSERS*D
 :||: | | | ||:   : |:|| | ||::|:|:| ||:|    :
QQKYIDNSEPGEGLLIFENVVLPFTNPIPHNTQLYKIMTTRLNEVAGV
                770       780       790
```

A related GBS gene <SEQ ID 8927> and protein <SEQ ID 8928> were also identified. Analysis of this protein sequence reveals the following:

```
This protein might be involved in vancomycin resistance
```

The protein has homology with the following sequences in the databases:

```
>GP|8100663|gb|AAF72347.1|AF192329_8|AF192329 TrsE-like protein
            {Enterococcus faecalis}

 Score =  427 bits (1086), Expect = e-118
 Identities = 257/785 (32%), Positives = 431/785 (54%), Gaps = 28/785 (3%)

Query: 9    DKKQKTKTQKQEIS-----------PSTVN-TLAYQGLFQNGLMQVSPSYFSQTYLLGDV 56
            +K + T+ Q++EI            P T   ++ Y+ ++ +G+ +VSP  FS+     D+
Sbjct: 11   EKTKLTRAQRKEIDAVIRKYKGDGRPHTAQQSIPYEVMYPDGVCRVSPGVFSKCIEFADI 70

Query: 57   NYQTVGLDDKGAIVEKYSDLINSLDDKTNFQLTIFNQKVNLEKFRKSILYPLQEDGFDTY 116
```

```
           +YQ    D + AI EK  DL N +D   + Q +  N+KV+  ++ KS     Q D FD
Sbjct: 71  SYQLAQPDTQTAIFEKLCDLYNYVDASIHIQFSFLNRKVDPVQYAKSFEIAPQGDDFDDI 130

Query: 117 RDELNRMMDANLEAGENNFSAVKFLSFGKSDQTPKLAFRSLSQIGEYFKSGFSEIDVSLG 176
           R E   ++   L  G N     K+L+F    ++ K A   L +IG      F  +
Sbjct: 131 RAEYTGILQKQLANGNNGMVKTKYLTFTIEAESVKAARARLKRIGFDLLGYFKSMGAVAH 190

Query: 177 LLGGEERVNVLADMLRGENHL-PFSYKDLTLSGQSTKHFIAPTYLSFKHKNHIELDDRLL 235
           ++ G ER+N+L  +   +  +  F +K L  SG STK FIAP+ L F +     +  +
Sbjct: 191 VMDGWERLNLLHGVYHPDGEIFNFDWKWLAPSGLSTKDFIAPSSLCFGNAKTFGMGGKYG 250

Query: 236 QIVYVRDYGMELGDKFIRDLMQSDLEVMISLHAKGSTKSETMTKLRTKKTLMESQKIGEQ 295
            + +++    EL D  + D + ++  V+++LH +   +++ +  ++ K T +++ KI EQ
Sbjct: 251 AVSFLQILSPELSDDMLADFLNTESGVLVNLHVQAIEQTKAIKTIKRKITDLDAMKIAEQ 310

Query: 296 QKMARTGIYLEKVGHVLENNIDEAEALLQTMTQTGDKLFDTVFLIGVLADTEDQLKQSLD 355
           +K  R+G  ++ +   L    ++A+ LL  +    ++LF   FL+  +ADT+ +L   +
Sbjct: 311 KKAVRSGYDMDILPSDLATYGEDAKKLLTKLQTRNERLFQLTFLVLNVADTKQKLNNDVF 370

Query: 356 IIKQVAGSNDMIIDNLTYMQEAAFNSLLPFGKNYLEGVSRSLLTSNIAVNAPWTSVDIHD 415
               VA  ++  +  L Y QE    S LP G N ++ + RSL TS++AV  P+ + ++
Sbjct: 371 QAAGVAQKHNCPLVRLDYQQEQGLASSLPLGVNQIK-IQRSLTTSSVAVFVPFVTQELFQ 429

Query: 416 KGGK-FYGINQISSNIISIDRGKLNTPSGLILGTSGAGKGMATKHEIISTKLKEADSDTE 474
            G   +YGIN  S N+I +DR +   P+ L LGT G+GK M+ K EI+S  L   D   +
Sbjct: 430 GGAAMYYGINAKSRNMIMLDRKQARCPNALKLGTPGSGKSMSCKSEIVSVFLTTPD---D 486

Query: 475 IIIVDPENEYSIIGQAFGGESIDIAPDSTTFLNVLELS-DENMDEDPVKVKSEFLLSWIG 533
           I I DPE EY  + +   G+ I ++P S  F+N L+++ + + D++P+ +KS+F+LS+
Sbjct: 487 IFISDPEAEYYPLVKRLHGQVIRLSPTSKDFVNPLDINLNYSEDDNPLALKSDFVLSFCE 546

Query: 534 KLLDRK--MDGREKSLIDRVTRLTYKHF-------DTPSLVEWVFVLSQQPEQEAKDLAL 584
            ++  K  ++  EK++IDR  R+ Y+ +       + P L +    L  Q   EA  +A
Sbjct: 547 LVMGGKNGLEAIEKTVIDRAVRVIYRPYLADPRPENMPILSDLHKALLDQHVPEADRVAQ 606

Query: 585 DMELYVEGSLDIFSHRTNIKTDSHFLIYNVKKLGDELKQIALMVIFDQIWNRVVKNQKLG 644
            ++LYV GSL++F+HRTN+   +  + +++K+LG +LK++ ++++ DQIW RV  N+  G
Sbjct: 607 ALDLYVSGSLNVFNHRTNVDIGNRLVSFDIKELGKQLKKLGMLIVQDQIWGRVTANRSQG 666

Query: 645 KKTWIYFDEMQLLLLDKYASDFFFKLWSRVRKYGAIPTGITQNVETLLLDANGRRIIANS 704
           K TW + DE  LLL ++  + +  ++W R RK+G IPTG TQNV+ LL       I+ NS
Sbjct: 667 KATWYFADEFHLLLKEEQTAAYSAEIWKRFRKWGGIPTGATQNVKDLLSSPEIENILENS 726

Query: 705 EFMILLKQAKSDREELVHMLGLSKELEKYLVNPEKGAGLIKAGSTVVPFKNKIPQHTKLF 764
           +F+ LL QA  DR+ L   L LS E +KY+ N E G GL+   + V+PF N IP +T+L+
Sbjct: 727 DFITLLNQASGDRKILAERLNLSTEQQKYIDNSEPGEGLLIFENVVLPFTNPIPHNTQLY 786

Query: 765 DIMST 769
            IM+T
Sbjct: 787 KIMTT 791
```

SEQ ID 8926 (GBS75) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 11; MW 89.8kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 20 (lane 6; MW 114.7kDa).

GBS75-GST was purified as shown in Figure 197, lane 8.

GBS329 was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 8; MW 89kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 174 (lane 2; MW 114kDa).

GBS329-GST was purified as shown in Figure 220, lanes 9 & 10.

Example 1891

A DNA sequence (GBSx1999) was identified in *S.agalactiae* <SEQ ID 5875> which encodes the amino acid sequence <SEQ ID 5876>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2442(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1892

A DNA sequence (GBSx2000) was identified in *S.agalactiae* <SEQ ID 5877> which encodes the amino acid sequence <SEQ ID 5878>. This protein is predicted to be DNA-directed RNA polymerase ii largest subunit. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4393(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1893

A DNA sequence (GBSx2001) was identified in *S.agalactiae* <SEQ ID 5879> which encodes the amino acid sequence <SEQ ID 5880>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.92   Transmembrane  256 - 272 ( 250 - 277)
   INTEGRAL    Likelihood = -8.28   Transmembrane  216 - 232 ( 213 - 244)
   INTEGRAL    Likelihood = -8.12   Transmembrane  151 - 167 ( 148 - 191)
   INTEGRAL    Likelihood = -7.27   Transmembrane   57 -  73 (  54 -  80)
   INTEGRAL    Likelihood = -6.74   Transmembrane   93 - 109 (  88 - 111)
   INTEGRAL    Likelihood = -3.50   Transmembrane  172 - 188 ( 168 - 191)
   INTEGRAL    Likelihood = -2.76   Transmembrane  113 - 129 ( 110 - 130)

----- Final Results -----
               bacterial membrane --- Certainty=0.4970(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG38039 GB:AF295925 Orf23 [Streptococcus pneumoniae]
 Identities = 71/86 (82%), Positives = 83/86 (95%)

Query: 37  VKSLADFNPTVWSYMTAITKGIMQPLGVAILAVVLVLEFSKMAKKIANSGGAMTFEAIAP 96
            +KSL+ +NPTVW+YM++ITK +MQPLGVAIL+VVL+LEFSKMAKKIANSGGAMTFEA+AP
Sbjct: 1   MKSLSSYNPTVWTYMSSITKSVMQPLGVAILSVVLILEFSKMAKKIANSGGAMTFEALAP 60

Query: 97  MIVSYIMVAVVITNTTVIVEAIIAIA 122
            M++SYIMVAVVITNTTVIVEAII IA
Sbjct: 61  MLISYIMVAVVITNTTVIVEAIIGIA 86
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1894

A DNA sequence (GBSx2002) was identified in *S.agalactiae* <SEQ ID 5881> which encodes the amino acid sequence <SEQ ID 5882>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.54   Transmembrane   32 -  48 (  25 -  52)
   INTEGRAL    Likelihood = -4.09   Transmembrane   63 -  79 (  62 -  80)

----- Final Results -----
               bacterial membrane --- Certainty=0.4015(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9933> which encodes amino acid sequence <SEQ ID 9934> was also identified. A related GBS nucleic acid sequence <SEQ ID 10777> which encodes amino acid sequence <SEQ ID 10778> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1895

A DNA sequence (GBSx2003) was identified in *S.agalactiae* <SEQ ID 5883> which encodes the amino acid sequence <SEQ ID 5884>. This protein is predicted to be TrsK-like protein (traK). Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -7.38   Transmembrane   66 -  82 (  62 -  85)

----- Final Results -----
               bacterial membrane --- Certainty=0.3951(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG38037 GB:AF295925 Orf21 [Streptococcus pneumoniae]
 Identities = 343/457 (75%), Positives = 385/457 (84%), Gaps = 24/457 (5%)
```

```
Query: 142 LIVIGGSGAGKTFRFVKPNLIQLNCSNIVVDPKDHLAEKTGKLFLENGYQVKVLDLVNMT 201
           + VIGGSG+GKTFRFVKPNLIQ+N SNIVVDPKDHLAEKTGKLFLE+GYQVKVLDLVNM
Sbjct: 1   MAVIGGSGSGKTFRFVKPNLIQMNSSNIVVDPKDHLAEKTGKLFLEHGYQVKVLDLVNMK 60

Query: 202 NSDGFNPFRYVETENDLNRMLTVYFNNTKGNGSRSDPFWDEASMTLVRAIASYLVDFYNP 261
           NSDGFNPFRY+ETENDLNRML VYFNNTKG+GSRSDPFWDEASMTLVRA+ASYLVDFYNP
Sbjct: 61  NSDGFNPFRYIETENDLNRMLAVYFNNTKGSGSRSDPFWDEASMTLVRALASYLVDFYNP 120

Query: 262 PGS---------------------SKQEQEARRKRGRYPAFSEIGKLIKLLSKGDNQDKS 300
           P +                      K+E E R+KRGR   F E    +    +   + KS
Sbjct: 121 PKTREQLIEESRLSQKEYQNLLKRQKKEVEERKKRGRLSKFCESQNSLNTYPRVKTR-KS 179

Query: 301 ILEVLFEDYAKKYGHENFTMRNWADFQNYKDKTLDSVIAVTTAKFALFNIQSVIDLTQRD 360
           +LE+LFE+YAKKYG ENFTMRNWADFQNYKDKTLDSVIAVTTAKFALFNIQSV+DLT+RD
Sbjct: 180 VLEILFENYAKKYGTENFTMRNWADFQNYKDKTLDSVIAVTTAKFALFNIQSVMDLTKRD 239

Query: 361 TMDLKTWGTQKTMVYLVIPDNDTTFRFLSAL-FFSTVFSTLTRQADV-DFKGQLPIHVRS 418
           T+D+KTWG +K+MVYLVIPDND+TFRFLSAL FF+  F T  + + +   + +LP+HVR
Sbjct: 240 TLDMKTWGQEKSMVYLVIPDNDSTFRFLSALLFFNPYFQTPNKTSQILMLRVRLPLHVRV 299

Query: 419 YLDEFANVGEIPDFAEQTSTVRSRNMSLVPILQNIAQLQGLYKEKEAWKTILGNCDSLLY 478
           YLDEFAN+GEIPDFAEQTSTVRSRNMSLVPILQNIAQLQGLYKEKEAWKTILGNCDSL+Y
Sbjct: 300 YLDEFANIGEIPDFAEQTSTVRSRNMSLVPILQNIAQLQGLYKEKEAWKTILGNCDSLVY 359

Query: 479 LGGNDEETFKFMSGLLGKQTVDVRSTSRSFGQTGSSSTSHQKIARDLMTADEVGTMKRDE 538
           LGGNDE+TFKFMSGLLGKQT+DVR+TSRSFGQTGS S SHQKIARDLMT DEVG MKR E
Sbjct: 360 LGGNDEDTFKFMSGLLGKQTIDVRNTSRSFGQTGSGSLSHQKIARDLMTPDEVGNMKRHE 419

Query: 539 CLVRIAGVPVFRTKKYFPLKHKHWKLLADKETDDRWW 575
           CLVRIA +PVF++KKY   KH +WK LA++ETD+R W
Sbjct: 420 CLVRIANMPVFKSKKYNSTKHPNWKYLANQETDERRW 456
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8929> and protein <SEQ ID 8930> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 5
McG: Discrim Score:      5.53
GvH: Signal Score (-7.5): -0.78
     Possible site: 34
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 1 value:  -7.38 threshold:  0.0
   INTEGRAL    Likelihood = -7.38   Transmembrane    66 -  82 (  62 -  85)
   PERIPHERAL  Likelihood =  1.75     338
 modified ALOM score:   1.98

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.3951(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
33.9/50.9% over 419aa
                                                          Lactococcus lactis
  GP|3582206| trsK protein (traK) Insert characterized
  PIR|T43089|T43089 transfer complex protein TrsK - plasmid pMRC01 Insert characterized

ORF00383(715 - 2004 of 2415)
```

```
GP|3582206|gb|AAC56002.1||AE001272(23 - 442 of 530) trsK protein (traK) {Lactococcus
lactis}PIR|T43089|T43089 transfer complex prote
in TrsK - Lactococcus lactis plasmid pMRC01
%Match = 10.1
%Identity = 33.8  %Similarity = 50.8
Matches = 141  Mismatches = 193  Conservative Sub.s = 71

519       549       579       609       639       669       699       729
SFLAFILGVLMMTLVYLYVSTGQKVYREGEEYGSARFGTSKEKRNFYSKNPFNDTILARDVRLTLLEKKKPQFDRNKNLI
                                                            |    :         |:|::
                                                     MNGTILGVLDNKIIYQDNTTKPNRNVM
                                                             10        20

759       789       816       846       876       906       936       966
VIGGSGAGKTFRFVKPNLIQLNCSNIVV-DPKDHLAEKTGKLFLENGYQVKVLDLVNMTNSDGFNPFRYVETENDLNRML
||||||: ||   |  ||     ::||| |||  | |||  : |  ||:| |::: || :|| :||| |:| :     :
VIGGSGSYKTQSVVITNLFNETKNSIVVTDPKGELYEKTAGIKLAQGYEVHVVNFANMAHSDRYNPFDYIERDIQAESVA
           40        50        60        70        80        90       100

996       1026      1056      1086      1116      1146      1176      1194
TVYFNNTKGNGSRSDPFWDEASMTLVRAIASYLVDFYNPPGSSKQEQEARRKRGRYPAFSEIGKLIKLLSKGD----NQD
|    :    | :    |      |::|:  :::                  : | |    :  :| :|   |    |:|
TKIVQSENAEGKK--DVWFSTQRQLLKALILFVM------------------KERSPEQRNLAGVINVLQTFDSEPINKD
          120         130                         140       150       160

1221      1251      1281      1311      1341      1371      1401      1431
K-SILEVLFEDYAKKYGHENFTMRNWADFQNYKDKTLDSVIAVTTAKFALFNIQSVIDLTQRDTMDLKTWGTQKTMVYLV
: | |: ||   | |  |          |:  |     |:|:   |  : |  : | ::|      |:  | :| ::|::
ENSDLDNLF--LALKITHPARIAYELG-FKKAKGDMKASIISSLLATISKFTDEEVSNFTSISDFHLQDIGRKKIVLYVI
            180       190        200       210       220       230       240

1461      1491      1521      1551      1581      1611      1641      1671
IPDNDTTFRFLSALFFSTVFSTLTRQADVDFKGQLPIHVRSYLDEFANVGEIPDFAEQTSTVRSRNMSLVPILQNIAQLQ
||  | |:  :  |||| :|  | : |  :   :||  |   |||| |:|: | : |  :| |   : :  | | : |||
IPVMDNTYESFINLFFSQMFDELYKLASSN-GAKLPQEVDFILDEFVNLGKFPKYEEFLATCRGYGIGVTTICQTLTQLQ
             260       270        280       290       300       310       320

1701      1731      1761      1791      1809      1839      1869      1899
GLYKEKEAWKTILGNCDSLLYLGGNDEETFKFMSGLLGKQTVDVR----STSRSFGQTGSSSTSHQKIARDLMTADEVGT
 ||  ||  ::||||    : |  ::| | |: | |||| || |     ||| |   : | | |:   :| ||| ||:
SLY-GKEKAESILGNHAVKICLNASNEATAKYFSELLGKSTVKVETGSESTSHSKETSTSKSDSYSYTSRQLMTPDEIIR
               340       350       360       370       380       390       400

1929      1956      1974      2004      2034      2064      2094      2124
MKRDECLVRIAGV-PVFRTK----KYFPLKHKHWKLLADKETDDRWWNYHINPLAKEEELDLSDYQIRDLSTETSLH**K
|   : |:      |:  ||    | ||      ||  :|       :          :  :     :::::
MPDTQSLLIFTNQKPIKATKAFQFKLFPDADSKVKLEQNKYVGITSKSQLEKYNDLSVKWEEKLQSLKNITVTEEEEKDL
               420       430       440       450       460       470       480
```

SEQ ID 5884 (GBS11d) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 151 (lane 6; MW 61kDa) and in Figure 182 (lane 10; MW 61kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 12 (lane 5; MW 91.5kDa).

Example 1896

A DNA sequence (GBSx2004) was identified in *S.agalactiae* <SEQ ID 5885> which encodes the amino acid sequence <SEQ ID 5886>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.4192(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9935> which encodes amino acid sequence <SEQ ID 9936> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1897

A DNA sequence (GBSx2005) was identified in *S.agalactiae* <SEQ ID 5887> which encodes the amino acid sequence <SEQ ID 5888>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3391(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1898

A DNA sequence (GBSx2006) was identified in *S.agalactiae* <SEQ ID 5889> which encodes the amino acid sequence <SEQ ID 5890>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.03   Transmembrane   68 -  84 (  64 -  90)
   INTEGRAL    Likelihood = -7.06   Transmembrane   33 -  49 (  27 -  50)
   INTEGRAL    Likelihood = -5.73   Transmembrane  106 - 122 ( 105 - 123)
   INTEGRAL    Likelihood = -4.46   Transmembrane    6 -  22 (   3 -  24)
   INTEGRAL    Likelihood = -2.13   Transmembrane  154 - 170 ( 154 - 170)
   INTEGRAL    Likelihood = -0.53   Transmembrane  180 - 196 ( 180 - 196)

----- Final Results -----
               bacterial membrane --- Certainty=0.5012(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9937> which encodes amino acid sequence <SEQ ID 9938> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA11325 GB:D78257 ORF8 [Enterococcus faecalis]
 Identities = 35/102 (34%), Positives = 57/102 (55%), Gaps = 4/102 (3%)

Query: 90  TRNQAVLVQVGKQVPPIIFLLFL-VNASILEEIVYRQLLWEKLTF--PFEQIGVTSFLFV 146
           T N + L+++    V P++ +L L + A I+EEIV+R  +    L         I ++SFLF
Sbjct: 7   TANDSTLIKLFSGVSPVLVVLLLGIAAPIMEEIVFRGGIIGYLVENNALLAILISSFLFG 66
```

```
Query: 147 LSHGPNQLGSWLIYSCLGLTLAVVRLKT-DCMTAIALHLLWN 187
           + HGP    S+ +Y  +G+ L+V   KT D   +I++H L N
Sbjct: 67  IIHGPTNFISFGMYFFMGIILSVSYYKTKDLRVSISIHFLNN 108
```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8931> and protein <SEQ ID 8932> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 4
McG: Discrim Score:      9.32
GvH: Signal Score (-7.5): -5.41
     Possible site: 45
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 6 value: -10.03 threshold:  0.0
   INTEGRAL    Likelihood =-10.03   Transmembrane   68 -  84 (  64 -  90)
   INTEGRAL    Likelihood = -7.06   Transmembrane   33 -  49 (  27 -  50)
   INTEGRAL    Likelihood = -5.73   Transmembrane  106 - 122 ( 105 - 123)
   INTEGRAL    Likelihood = -4.46   Transmembrane    6 -  22 (   3 -  24)
   INTEGRAL    Likelihood = -2.13   Transmembrane  154 - 170 ( 154 - 170)
   INTEGRAL    Likelihood = -0.53   Transmembrane  180 - 196 ( 180 - 196)
   PERIPHERAL  Likelihood =  1.38     131
 modified ALOM score:   2.51

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.5012(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01326(568 - 861 of 1188)
EGAD|148261|158156(7 - 108 of 120) hypothetical protein {Enterococcus faecalis}
GP|1402529|dbj|BAA11325.1||D78257 ORF8 {Enterococcus faecalis}
%Match = 5.9
%Identity = 34.7  %Similarity = 60.4
Matches = 35  Mismatches = 37  Conservative Sub.s = 26

303       333       363       393       423       453       483       513
Y*L*RFI*EVTMIRIVLFYLAIQLNGLLVSLFLKEYLTIEGIVLLQLVLLSVTCLEIARHKTVPLKIVGVQNRLSWLLLG

543       573       603       633       660       690       714       744
FVAMVAFAVFISFLFPVQTRNQAVLVQVGKQVPPIIFLLFL-VNASILEEIVYRQLLWEKLT--FPFEQIGVTSFLFVLS
                  | | : |:::   | |:: :|:| : | |:||||:|  :   |     :  | ::|||| :
            MQGHTTTANDSTLIKLFSGVSPVLVVLLLGIAAPIMEEIVFRGGIIGYLVENNALLAILISSFLFGII
                    10        20        30        40        50        60

774       804       831       861       891       921       951       981
HGPNQLGSWLIYSCLGLTLAVVRLKT-DCMTAIALHLLWNSLAYVVTFL*YQNQECFRIMEAPYV**GIEKRGGHYVI*T
|||  : |: :|  :|: |:|   || |   :|::|:| |
HGPTNFISFGMYFFMGIILSVSYYKTKDLRVSISIHFLNNLFPAIAIAYGLI
          80        90       100       110       120
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1899

A DNA sequence (GBSx2007) was identified in *S.agalactiae* <SEQ ID 5891> which encodes the amino acid sequence <SEQ ID 5892>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2490(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9939> which encodes amino acid sequence <SEQ ID 9940> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1900

A DNA sequence (GBSx2008) was identified in *S.agalactiae* <SEQ ID 5893> which encodes the amino acid sequence <SEQ ID 5894>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5298(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC98423 GB:L29323 unknown [Streptococcus pneumoniae]
 Identities = 68/126 (53%), Positives = 88/126 (68%)

Query: 1   MNLLHKKSILDCTELEERIHQAETNQLLQKILSLPNFDCDFEVTFEDDYHKEMNDPLFYE 60
           M  L+K+SILDC ELE  +H AE  QL ++I  +PN+ C+FEVTF DDYHK+ N PLFYE
Sbjct: 1   MKALNKESILDCDELETELHDAEIKQLDEQIFLMPNYPCEFEVTFLDDYHKKHNYPLFYE 60

Query: 61  SNLHQISDFMETRDIKNGVDTLLTKDNHLAFRAFGENYSARGKEGILTTLVTVKCFGEGR 120
           S L  I +F+E++DIKNG D  +    +L F  +G+ Y A GKEGILTT VTVK F E +
Sbjct: 61  SYLQNIMEFLESQDIKNGADAFVDDHQNLVFVLYGQGYRAEGKEGILTTQVTVKAFDEDK 120

Query: 121 MPIDMS 126
            PI+ +
Sbjct: 121 KPINFA 126
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1901

A DNA sequence (GBSx2009) was identified in *S.agalactiae* <SEQ ID 5895> which encodes the amino acid sequence <SEQ ID 5896>. This protein is predicted to be methyl transferase. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
                        bacterial cytoplasm --- Certainty=0.1209(Affirmative) < succ>
                         bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC98421 GB:L29323 methyl transferase [Streptococcus pneumoniae]
 Identities = 323/449 (71%), Positives = 389/449 (85%), Gaps = 3/449 (0%)

Query: 1   MKFLDLFAGIGGFRLGMESQGHKCLGFCEIDKFARTSYKAMFNTEGEIEYHDIKEVTDHD 60
           M+F+DLF+GIGGFRLGMES GH+C+GFCEIDKFAR SYK++F TEGEIE+HDI++V+D +
Sbjct: 1   MRFIDLFSGIGGFRLGMESVGHECIGFCEIDKFARESYKSIFQTEGEIEFHDIRDVSDDE 60

Query: 61  FRQFRGQVDIICGGFPCQAFSLAGRRLGFEDTRGTLFFEIARAAKQIQPRFLFLENVKGL 120
           F++ RG+VD+ICGGFPCQAFS+AGRRLGFEDTRGTLFFEIARAAKQIQPRFLFLENVKGL
Sbjct: 61  FKKLRGKVDVICGGFPCQAFSIAGRRLGFEDTRGTLFFEIARAAKQIQPRFLFLENVKGL 120

Query: 121 LNHDEGRTFATILSTLDELGYDVEWQVLNSKDFQVPQNRERVFIIGHSRRYRSRFIFPLR 180
           LNHD+GRTF TIL+TLDELG+DVEWQ+LNSKDF VPQNRERVFIIGHSR+  +R  FP R
Sbjct: 121 LNHDKGRTFTTILTTLDELGFDVEWQMLNSKDFGVPQNRERVFIIGHSRKRGTRLGFPFR 180

Query: 181 RED---SPAHLERLGNINPSKHGLNGEVYLTSGLAPTLTRGKGEGAKIAIPVLTPDRLEK 237
           RE    +P  L+ LGN+NPSK G++G+VY + GLAPTL RGKGEG KIAIP +TPDRL+K
Sbjct: 181 REGQATNPETLKILGNLNPSKSGMSGKVYYSEGLAPTLVRGKGEGFKIAIPCMTPDRLDK 240

Query: 238 RQHGRRFKDNQDPMFTLTSQDKHGVVVAGNLPTSFDQTGRVFDISGLSPTLTTMQGGDKV 297
           RQ+GRRFKDNQ+PMFTL +QD+HG+VV G+LPTSF +TGRV+   GLSPTLTTMQGGDK+
Sbjct: 241 RQNGRRFKDNQEPMFTLNTQDRHGIVVVGDLPTSFKETGRVYGSEGLSPTLTTMQGGDKI 300

Query: 298 PKILLREELPFLKIKEATKTGYAKATLGDSVNLAYPDSTKRRGRVGKGISNTLTTSDNMG 357
           PKIL+ E + FLK++EATK GYA+A +GDS+NL  P S  RRGRVGKGI+NTLTTS  MG
Sbjct: 301 PKILIPEPIQFLKVREATKKGYAQAEIGDSINLERPSSQHRRGRVGKGIANTLTTSGQMG 360

Query: 358 VVVAALEYRQDKWYEVTGIVLEGKLYRLRIRRLTPRECFRLQGFPDWAYERAESVSSKSQ 417
           VVVA+ E    + Y+V G++++G+ YRLRIRR+TP+ECFRLQGFPDWA+E A  VSS SQ
Sbjct: 361 VVVASYEGEDKQVYQVAGVLIDGQFYRLRIRRITPKECFRLQGFPDWAFEAARKVSSNSQ 420

Query: 418 LYKQAGNSVTVTVIEAIAREFRRTEEEEK 446
           LYKQAGNSVTV VI AIA++ +  EE+++
Sbjct: 421 LYKQAGNSVTVPVIAAIAKKLKEVEEKDE 449
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2435> which encodes the amino acid sequence <SEQ ID 2436>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
                        bacterial cytoplasm --- Certainty=0.1725(Affirmative) < succ>
                         bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 60/75 (80%), Positives = 69/75 (92%)

Query: 1  MKFLDLFAGIGGFRLGMESQGHKCLGFCEIDKFARTSYKAMFNTEGEIEYHDIKEVTDHD 60
          MKFLDLFAGIGGFRLG+ +Q H+C+GFCEIDKFAR SYKA++ TEGEIE+HDI++VTD D
Sbjct: 4  MKFLDLFAGIGGFRLGLINQCHECIGFCEIDKFARQSYKAIYETEGEIEFHDIRQVTDQD 63

Query: 61 FRQFRGQVDIICGGF 75
          FRQ RGQVDIICGGF
Sbjct: 64 FRQLRGQVDIICGGF 78
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1902

A DNA sequence (GBSx2010) was identified in *S.agalactiae* <SEQ ID 5897> which encodes the amino acid sequence <SEQ ID 5898>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.71    Transmembrane    8 -  24 (   3 -  30)

----- Final Results -----
               bacterial membrane --- Certainty=0.4885(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9941> which encodes amino acid sequence <SEQ ID 9942> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5899> which encodes the amino acid sequence <SEQ ID 5900>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.81    Transmembrane   20 -  36 (  19 -  36)

----- Final Results -----
               bacterial membrane --- Certainty=0.1723(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 16/33 (48%), Positives = 26/33 (78%)

Query: 1  MNKMIWWILGGIYLISIIILIVEIIRAPEMDDH 33
           ++KM WW+L G++ +  I LI+E+I APEM+D+
Sbjct: 12 VSKMFWWLLLGVWGLRTIWLIIEVITAPEMEDY 44
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1903

A DNA sequence (GBSx2011) was identified in *S.agalactiae* <SEQ ID 5901> which encodes the amino acid sequence <SEQ ID 5902>. This protein is predicted to be ifn-response binding factor 1 (irebf-1). Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4771(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD41248 GB:AF106927 unknown [Streptococcus suis]
 Identities = 258/272 (94%), Positives = 266/272 (96%)

Query: 1   MKRITANQYQTSERYYKLPKILFESERYKDMKLEVKVAYAVLKDRLELSLSKGWIDEDGA 60
            MKRITANQYQTSERYYKLPKILFESERYKDMKLEVKVAYAVLKDRLELSLSKGWIDEDGA
```

```
Sbjct: 1   MKRITANQYQTSERYYKLPKILFESERYKDMKLEVKVAYAVLKDRLELSLSKGWIDEDGA 60

Query: 61  IYLIYSNSNLMALLGCSKSKLLSIKKTLREYGLIDEVQQSSSERGRMANKIYLGELEHEP 120
            IYLIYSNSNLMALLGCSKSKLLSIKKTLREYGLIDEVQQSSSE+GRMANKIYLGELEHE
Sbjct: 61  IYLIYSNSNLMALLGCSKSKLLSIKKTLREYGLIDEVQQSSSEKGRMANKIYLGELEHET 120

Query: 121 TPVLHTDGASVKKTLGESQRKTGPVLYSAPSETEGSETKYSETEGSDLVMKDEEERQLVD 180
            TPVLHTDGASVKKTLG SQRKTGPVL SAPSETEGSETKYSET+GSD +++DEEERQ VD
Sbjct: 121 TPVLHTDGASVKKTLGGSQRKTGPVLNSAPSETEGSETKYSETKGSDFLIEDEEERQQVD 180

Query: 181 EKKEENFTSKVDGVTKYDRDYIWGLVHDQLRQTGLSQSASDYAMIYFSDRYQYALEQMRF 240
            EK+EENFTSKVDGVT+YDRDYIWGLVHDQLRQTGLSQSASDYAMIYFSDRYQYALE MRF
Sbjct: 181 EKQEENFTSKVDGVTRYDRDYIWGLVHDQLRQTGLSQSASDYAMIYFSDRYQYALEHMRF 240

Query: 241 ARSAEVIAEYVFNGVLSEWTKQLRRQEVKGGE 272
            ARSAEVIAEYVFNGVLSEWTKQLRRQEVKGG+
Sbjct: 241 ARSAEVIAEYVFNGVLSEWTKQLRRQEVKGGD 272
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5903> which encodes the amino acid sequence <SEQ ID 5904>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5248(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 84/122 (68%), Positives = 99/122 (80%), Gaps = 2/122 (1%)

Query: 145 VLYSAPSETEGSETKYSETEGSDLVMKDEEERQLVD--EKKEENFTSKVDGVTKYDRDYI 202
            VL SAPSETE SET+ SET+ S+LV++DEEER+     +K E +FT +VD VTKYD+DYI
Sbjct: 1   VLNSAPSETEKSETEGSETKESNLVIEDEEERKECTSVKKTEGHFTRQVDQVTKYDKDYI 60

Query: 203 WGLVHDQLRQTGLSQSASDYAMIYFSDRYQYALEQMRFARSAEVIAEYVFNGVLSEWTKQ 262
            W LVH QLR+ GLSQ+ASD  M YF +RY YALE +RFAR+AE IAEYVFNGVLSEWTKQ
Sbjct: 61  WSLVHSQLREGGLSQAASDLVMSYFEERYAYALEHIRFARTAEAIAEYVFNGVLSEWTKQ 120

Query: 263 LR 264
            LR
Sbjct: 121 LR 122
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1904

A DNA sequence (GBSx2012) was identified in *S.agalactiae* <SEQ ID 5905> which encodes the amino acid sequence <SEQ ID 5906>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4191(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9375> which encodes amino acid sequence <SEQ ID 9376> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1905

A DNA sequence (GBSx2013) was identified in *S.agalactiae* <SEQ ID 5907> which encodes the amino acid sequence <SEQ ID 5908>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3723(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1906

A DNA sequence (GBSx2014) was identified in *S.agalactiae* <SEQ ID 5909> which encodes the amino acid sequence <SEQ ID 5910>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3053(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1907

A DNA sequence (GBSx2015) was identified in *S.agalactiae* <SEQ ID 5911> which encodes the amino acid sequence <SEQ ID 5912>. This protein is predicted to be 50S ribosomal protein L7/l12 (rplL). Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1034(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9943> which encodes amino acid sequence <SEQ ID 9944> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB11881 GB:Z99104 ribosomal protein L12 (BL9) [Bacillus subtilis]
 Identities = 83/123 (67%), Positives = 95/123 (76%), Gaps = 2/123 (1%)

Query: 6   MALNIENIIAEIKEATILELNDLVKAIEEEFGVTAAAPVAAA--AAGGEAAAAKDSFDVE 63
            MALNIE IIA +KEAT+LELNDLVKAIEEEFGVTAAAPVA A  AA G AA  +  FD+
Sbjct: 1   MALNIEEIIASVKEATVLELNDLVKAIEEEFGVTAAAPVAVAGGAAAGGAAEEQSEFDLI 60

Query: 64  LTAAGDKKVGVIKVVREITGEGLKEAKAIVDNAPSVIKEGASEAEANEIKEKLEAAGASV 123
            L  AG +K+ VIKVVREITG GLKEAK +VDN P  +KEG ++ EA E+K KLE  GASV
Sbjct: 61  LAGAGSQKIKVIKVVREITGLGLKEAKELVDNTPKPLKEGIAKEEAEELKAKLEEVGASV 120

Query: 124 TLK 126
             +K
Sbjct: 121 EVK 123
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5913> which encodes the amino acid sequence <SEQ ID 5914>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1164(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 104/126 (82%), Positives = 113/126 (89%)

Query: 1   MEEITMALNIENIIAEIKEATILELNDLVKAIEEEFGVTAAAPVAAAAAGGEAAAAKDSF 60
            +EEITMALNIENIIAEIKEA+ILELNDLVKAIEEEFGVTAAAPVAAAAAGG   AAKDSF
Sbjct: 1   LEEITMALNIENIIAEIKEASILELNDLVKAIEEEFGVTAAAPVAAAAAGGAEEAAKDSF 60

Query: 61  DVELTAAGDKKVGVIKVVREITGEGLKEAKAIVDNAPSVIKEGASEAEANEIKEKLEAAG 120
            DVELT+AGDKKVGVIK VREITG GLKEAK +VD AP+ +KEG + AEA EIK KLE AG
Sbjct: 61  DVELTSAGDKKVGVIKAVREITGLGLKEAKGLVDGAPANVKEGVAAAEAEEIKAKLEEAG 120

Query: 121 ASVTLK 126
            A++TLK
Sbjct: 121 ATITLK 126
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1908

A DNA sequence (GBSx2017) was identified in *S.agalactiae* <SEQ ID 5915> which encodes the amino acid sequence <SEQ ID 5916>. This protein is predicted to be ribosomal protein L10 (rplJ). Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1251(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
                    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB11880 GB:Z99104 ribosomal protein L10 (BL5) [Bacillus subtilis]
 Identities = 96/164 (58%), Positives = 125/164 (75%), Gaps = 1/164 (0%)

Query: 14  MSEAIIAKKAEQVELIAEKMKAAASIVVVDSRGLTVEQDTNLRRSLRESDVEFKVIKNSI 73
            MS AI  KK   VE IA K+K + S ++VD RGL V + T LR+ LRE++VE KV KN++
Sbjct: 1   MSSAIETKKVV-VEEIASKLKESKSTIIVDYRGLNVSEVTELRKQLREANVESKVYKNTM 59

Query: 74  LTRAAEKAGLEDLKELFVGPSAVAFSNEDVIAPAKVISDFAKDAEALEIKGGSVDGKFTS 133
              RA E+A L  L +   GP+A+AFS EDV+APAKV++DFAK+ EALEIK G ++GK ++
Sbjct: 60  TRRAVEQAELNGLNDFLTGPNAIAFSTEDVVAPAKVLNDFAKNHEALEIKAGVIEGKVST 119

Query: 134 VEEINALAKLPNKEGMLSMLLSVLQAPVRNVAYAVKAVAEKDEE 177
            VEE+ ALA+LP +EG+LSMLLSVL+APVRN+A A KAVAE+ EE
Sbjct: 120 VEEVKALAELPPREGLLSMLLSVLKAPVRNLALAAKAVAEQKEE 163
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5917> which encodes the amino acid sequence <SEQ ID 5918>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -5.47    Transmembrane     7 -  23 (   5 -  24)

----- Final Results -----
                   bacterial membrane --- Certainty=0.3187(Affirmative) < succ>
                    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 149/176 (84%), Positives = 162/176 (91%)

Query: 4   SQKIKTEVKLMSEAIIAKKAEQVELIAEKMKAAASIVVVDSRGLTVEQDTNLRRSLRESD 63
            S KIKTEVKLMSEAIIAKKAEQVELIAEKMKAAASIV+VDSRGLTV+QDT LRRSLRES
Sbjct: 23  SPKIKTEVKLMSEAIIAKKAEQVELIAEKMKAAASIVIVDSRGLTVDQDTVLRRSLRESG 82

Query: 64  VEFKVIKNSILTRAAEKAGLEDLKELFVGPSAVAFSNEDVIAPAKVISDFAKDAEALEIK 123
            VEFKVIKNSILTRAAEKAGL++LK++FVGPSAVAFSNEDVIAPAKVI+DF K A+ALEIK
Sbjct: 83  VEFKVIKNSILTRAAEKAGLDELKDVFVGPSAVAFSNEDVIAPAKVINDFTKTADALEIK 142

Query: 124 GGSVDGKFTSVEEINALAKLPNKEGMLSMLLSVLQAPVRNVAYAVKAVAEKDEEVA 179
            GG+++G  +S EEI ALA LPN+EGMLSMLLSVLQAPVRNVAYAVKAVAE  E  A
Sbjct: 143 GGAIEGAVSSKEEIQALATLPNREGMLSMLLSVLQAPVRNVAYAVKAVAENKEGAA 198
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1909

A DNA sequence (GBSx2018) was identified in *S.agalactiae* <SEQ ID 5919> which encodes the amino acid sequence <SEQ ID 5920>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.22    Transmembrane  125 - 141 ( 106 - 143)
   INTEGRAL    Likelihood = -1.91    Transmembrane  108 - 124 ( 106 - 124)

----- Final Results -----
                   bacterial membrane --- Certainty=0.3888(Affirmative) < succ>
                    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10931> which encodes amino acid sequence <SEQ ID 10932> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1910

A DNA sequence (GBSx2019) was identified in *S.agalactiae* <SEQ ID 5921> which encodes the amino acid sequence <SEQ ID 5922>. This protein is predicted to be Clp-like ATP-dependent protease binding subunit (clpC). Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3483(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA68910 GB:L34677 Clp-like ATP-dependent protease binding
           subunit [Bos taurus]
 Identities = 437/589 (74%), Positives = 514/589 (87%), Gaps = 5/589 (0%)

Query: 10  DPFGN-MDDIFNSLMGNMGGYNSENKRYLINGREVTPEEFSQYRQTGKLPGQELNNQNTP 68
            DPF N MDD+FN LMG M G NSEN+RYLINGREVTPEE++ +RQTGKLPG     Q
Sbjct: 2   DPFNNDMDDLFNQLMGGMNGVNSENRRYLINGREVTPEEYAAFRQTGKLPGVTDPTQ-AK 60

Query: 69  TNQVSADSVLTKLGTNLTDQARQHLLDPVIGRNKEIQETAEILARRTKNNPVLVGDAGVG 128
            T Q   DS+L KLG NLT +A++  LDPVIGRNKEIQETAEIL+RRTKNNPVLVGDAGVG
Sbjct: 61  TKQPQPDSMLAKLGRNLTQEAKEGKLDPVIGRNKEIQETAEILSRRTKNNPVLVGDAGVG 120

Query: 129 KTAVIEGLAQAIINGDVPAAIKNKEIISIDISSLEAGTQYRGSFEENIQNIIKEVKETGN 188
            KTAV+EGLAQAI+ GDVPAAIKNK+IISIDISSLEAGTQYRGSFEEN+Q +I EVK+ GN
Sbjct: 121 KTAVVEGLAQAIVAGDVPAAIKNKQIISIDISSLEAGTQYRGSFEENMQKLIDEVKKDGN 180

Query: 189 IILFFDEIHQILGAGSTGGDSGSKGLADILKPALSRGELTVIGATTQDEYRNTILKNAAL 248
            +ILFFDEIHQI+GAG+ G  SGSKG+ADILKPALSRGE+T+IGATTQDEYRNTILK+AAL
Sbjct: 181 VILFFDEIHQIIGAGNAGDASGSKGMADILKPALSRGEVTLIGATTQDEYRNTILKDAAL 240

Query: 249 ARRFNEVKVNAPSAQDTFNILMGIRNLYEQHHNVVLPDSVLKAAVDLSIQYIPQRSLPDK 308
            +RRFN+V VNAPS +DTF IL G+R LYE+HHNV LPD VLKAA+D S+QYIPQRSLPDK
Sbjct: 241 SRRFNQVTVNAPSKEDTFKILQGLRKLYEKHHNVSLPDEVLKAAIDYSVQYIPQRSLPDK 300

Query: 309 AIDLIDMTAAHLAAQHPVTDLKSLEKEIAEQRDKQEKAVNTEDFEEALKVKTRIEELQNQ 368
            AIDLID+TAAHLA++HPV D K++E+EI +   KQ++AV  ED++ A + K ++ +LQ+Q
Sbjct: 301 AIDLIDVTAAHLASKHPVKDAKTIEEEIKKTEAKQQEAVEKEDYQAAQEAKDQVAKLQDQ 360

Query: 369 IDNHTEGQKVTATINDIAMSIERLTGVPVSNMGASDIERLKELGNRLKGKVIGQNDAVEA 428
            + +H+E ++V AT +D+A ++ER+TG+PVS MGASDIERLK L  RL+GKVIGQ +AVEA
Sbjct: 361 LKDHSESERVVATPSDVAAAVERMTGIPVSKMGASDIERLKGLATRLEGKVIGQQEAVEA 420

Query: 429 VARAIRRNRAGFDDGNRPIGSFLFVGPTGVGKTELAKQLAFDMFGSKDAIVRLDMSEYND 488
            V+RAIRRNRAGFD+GNRPIGSFLFVGPTGVGKTELAKQLA DMFGS + I+RLDMSEY D
Sbjct: 421 VSRAIRRNRAGFDEGNRPIGSFLFVGPTGVGKTELAKQLALDMFGSTNDIIRLDMSEYTD 480

Query: 489 RTAVSKLIGATAGYVGYDDNSNTLTERIRRNPYSIVLLDEIEKADPQVITLLLQVLDDGR 548
            RTAVSKLIG TAGYVGYDDNSNTLTE++RR+PYSIVLLDEIEKA+PQVITLLLQVLDDGR
```

```
Sbjct: 481 RTAVSKLIGTTAGYVGYDDNSNTLTEKVRRHPYSIVLLDEIEKANPQVITLLLQVLDDGR 540

Query: 549 LTDGQGNTINFKNTVIIATSNAGFGNEAFTGDSDKDLKIMERISPYFRP 597
           LTDGQGNT++FKNT+IIATSNAGF ++A  G+   D K+M+++ PYFRP
Sbjct: 541 LTDGQGNTVDFKNTIIIATSNAGFSSDAVAGE---DAKLMDKLQPYFRP 586
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5923> which encodes the amino acid sequence <SEQ ID 5924>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2718(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 551/697 (79%), Positives = 616/697 (88%), Gaps = 3/697 (0%)

Query: 5   NFYNRDPFGNMDDIFNSLMGNMGGYNSENKRYLINGREVTPEEFSQYRQTGKLPGQELNN 64
           +F  +DPF NMDDIFN LM NMGGY SEN RYL+NGRE+TPEEF  YRQTG+LP
Sbjct: 3   HFSGKDPFVNMDDIFNQLMANMGGYRSENPRYLVNGREITPEEFQHYRQTGQLPVATTKA 62

Query: 65  QNTPTNQVSADSVLTKLGTNLTDQARQHLLDPVIGRNKEIQETAEILARRTKNNPVLVGD 124
            N+      ADSVLT+LGTNLT +ARQ  LDPVIGRNKEIQ+TAEILARRTKNNPVLVGD
Sbjct: 63  TNSQMLTPKADSVLTQLGTNLTQEARQGHLDPVIGRNKEIQDTAEILARRTKNNPVLVGD 122

Query: 125 AGVGKTAVIEGLAQAIINGDVPAAIKNKEIISIDISSLEAGTQYRGSFEENIQNIIKEVK 184
           AGVGKTAVIEGLAQAI+NGDVPAAIKNKEI+SIDISSLEAGTQYRGSFEE IQN+I+EVK
Sbjct: 123 AGVGKTAVIEGLAQAIVNGDVPAAIKNKEIVSIDISSLEAGTQYRGSFEETIQNLIQEVK 182

Query: 185 ETGNIILFFDEIHQILGAGSTGGDSGSKGLADILKPALSRGELTVIGATTQDEYRNTILK 244
           E GNIILFFDEIHQI+GAG+T  DSGSKGLADILKPALSRGELT+IGATTQDEYRNTILK
Sbjct: 183 EAGNIILFFDEIHQIVGAGATSSDSGSKGLADILKPALSRGELTLIGATTQDEYRNTILK 242

Query: 245 NAALARRFNEVKVNAPSAQDTFNILMGIRNLYEQHHNVVLPDSVLKAAVDLSIQYIPQRS 304
           NAALARRFNEVKVNAPSA+DTF+ILMGIRNLYEQHH++ LPD+VLKAAVD SIQYIPQRS
Sbjct: 243 NAALARRFNEVKVNAPSAEDTFHILMGIRNLYEQHHHITLPDNVLKAAVDYSIQYIPQRS 302

Query: 305 LPDKAIDLIDMTAAHLAAQHPVTDLKSLEKEIAEQRDKQEKAVNTEDFEEALKVKTRIEE 364
           LPDKAIDL+DMTAAHLAAQHPVTDLK+LE EIA+Q++ QEKAV  EDFE+AL  KTRIE
Sbjct: 303 LPDKAIDLLDMTAAHLAAQHPVTDLKTLETEIAKQKESQEKAVAKEDFEKALAAKTRIET 362

Query: 365 LQNQIDNHTEGQKVTATINDIAMSIERLTGVPVSNMGASDIERLKELGNRLKGKVIGQND 424
           LQ QI+ H + Q VTAT+NDIA S+ERLTG+PVSNMG +D+ERLK + +RLK  VIGQ++
Sbjct: 363 LQKQIEQHNQSQNVTATVNDIAESVERLTGIPVSNMGTNDLERLKGISSRLKSHVIGQDE 422

Query: 425 AVEAVARAIRRNRAGFDDGNRPIGSFLFVGPTGVGKTELAKQLAFDMFGSKDAIVRLDMS 484
           AV AVARAIRRNRAGFDDG RPIGSFLFVGPTGVGKTELAKQLA D+FGSKDAI+RLDMS
Sbjct: 423 AVAAVARAIRRNRAGFDDGKRPIGSFLFVGPTGVGKTELAKQLALDLFGSKDAIIRLDMS 482

Query: 485 EYNDRTAVSKLIGATAGYVGYDDNSNTLTERIRRNPYSIVLLDEIEKADPQVITLLLQVL 544
           EYNDRTAVSKLIG TAGYVGYDDN+NTLTER+RRNPY+IVLLDEIEKADPQ+ITLLLQVL
Sbjct: 483 EYNDRTAVSKLIGTTAGYVGYDDNNNTLTERVRRNPYAIVLLDEIEKADPQIITLLLQVL 542

Query: 545 DDGRLTDGQGNTINFKNTVIIATSNAGFGNEAFTGDSDKDLKIMERISPYFRPEFLNRFN 604
           DDGRLTDGQGNTINFKNTVIIATSNAGFG +        +  IM+RI+PYFRPEFLNRFN
Sbjct: 543 DDGRLTDGQGNTINFKNTVIIATSNAGFGQQ---DTETSESNIMDRIAPYFRPEFLNRFN 599

Query: 605 GVIEFSHLSKDDLSEIVDLMLDEVNQTIGKKGIDLVVDENVKSHLIELGYDEAMGVRPLR 664
            +I+F+HL K+ L EIVDLML EVNQT  KKGI L + ++ K+HLI+LGY+ AMG RPLR
Sbjct: 600 SIIKFNHLQKESLEEIVDLMLAEVNQTTAKKGISLTITDDAKAHLIDLGYNHAMGARPLR 659

Query: 665 RVIEQEIRDRITDYYLDHTDVKHLKANLQDGQIVISE 701
           R+IEQEIRDRITDYYLDH +VK L+A L++GQ+VI +
```

```
Sbjct: 660 RIIEQEIRDRITDYYLDHPEVKKLQAILKEGQLVIRQ 696
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1911

A DNA sequence (GBSx2020) was identified in *S.agalactiae* <SEQ ID 5925> which encodes the amino acid sequence <SEQ ID 5926>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -4.78    Transmembrane    8 -  24 (   7 -  25)

----- Final Results -----
               bacterial membrane --- Certainty=0.2911(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9945> which encodes amino acid sequence <SEQ ID 9946> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC73364 GB:AE000134 putative enzyme [Escherichia coli K12]
 Identities = 142/307 (46%), Positives = 195/307 (63%), Gaps = 6/307 (1%)

Query: 39  KELLESKKTLILHGALGTELESRGCDVSGKLWSAKYLIEDPAAIQTIHEDYIRAGADIVT 98
            + LL+ +  L+L GA+ TELE+RGC+++  LWSAK L+E+P  I+ +H DY RAGA
Sbjct: 8   RALLDKQDILLLDGAMATELEARGCNLADSLWSAKVLVENPELIREVHLDYYRAGAQCAI 67

Query: 99  TSTYQATLQGLAQVGVSESQTEDLIRLTVQLAKAAREQVWKSLTKEEKSERIYPLISGDV 158
            T++YQAT  G A  G+ E+Q++ LI  +V+LA+ ARE     L +  ++  +  L++G V
Sbjct: 68  TASYQATPAGFAARGLDEAQSKALIGKSVELARKAREAY---LAENPQAGTL--LVAGSV 122

Query: 159 GPYAAFLADGSEYTGLYDIDKQGLKNFHRHRIELLLDEGVDILALETIPNAQEAEALIEL 218
            GPY A+LADGSEY G Y    +  + FHR R+E LLD G D+LA ET+PN  E EAL EL
Sbjct: 123 GPYGAYLADGSEYRGDYHCSVEAFQAFHRPRVEALLDAGADLLACETLPNFSEIEALAEL 182

Query: 219 LAEDFPQVEAYMSFTSQDGKTISDGSAVADLAKAIDVSPQVVALGINCSSPSLVADFLQA 278
            L   +P+  A+ SFT +D + +SDG+ + D+   +   PQVVALGINC +       LQ
Sbjct: 183 LTA-YPRARAWFSFTLRDSEHLSDGTPLRDVVALLAGYPQVVALGINCIALENTTAALQH 241

Query: 279 IAEQTNKPLVTYPNSGEVYDGASQSWQSSPDHSHTLLENTSDWQKLGAQVVGGCCRTRPA 338
            +   T  PLV YPNSGE YD  S++W    +H   L +    WQ  GA+++GGCCRT PA
Sbjct: 242 LHGLTVLPLVVYPNSGEHYDAVSKTWHHHGEHCAQLADYLPQWQAAGARLIGGCCRTTPA 301

Query: 339 DIADLSA 345
            DIA L A
Sbjct: 302 DIAALKA 308
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8933> and protein <SEQ ID 8934> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 5
McG: Discrim Score:      5.48
GvH: Signal Score (-7.5): -2.64
     Possible site: 20
```

```
>>> Seems to have an uncleavable N-term signal seq
ALOM program  count: 1 value: -4.78 threshold:  0.0
   INTEGRAL    Likelihood = -4.78   Transmembrane    8 -  24 (   7 -  25)
   PERIPHERAL  Likelihood =  2.49     259
 modified ALOM score:   1.46

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.2911(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01312(412 - 1338 of 1644)
OMNI|NT01EC0303(55 - 357 of 358) conserved hypothetical protein
%Match = 23.8
%Identity = 46.6  %Similarity = 64.3
Matches = 142  Mismatches = 107  Conservative Sub.s = 54

288       318       348       378       408       438       468       498
LISQSFCS*FRL*GLLGIAHNVLGFTSVFHLLFSAIFITNYVTRNGDLMGRFKELLESKKTLILHGALGTELESRGCDVS
                                                   :: ||: :  |:| ||: ||||:|||:::
AWWPVLGWHSIQRRELRCGAGYRLLRCAMVLISLLNPETQNRSQNMSQNNPLRALLDKQDILLLDGAMATELEARGCNLA
               20        30        40        50        60        70        80

528       558       588       618       648       678       708       738
GKLWSAKYLIEDPAAIQTIHEDYIRAGADIVTTSTYQATLQGLAQVGVSESQTEDLIRLTVQLAKAAREQVWKSLTKEEK
  ||||| |:|:|  |: :| || ||||    |::||||  |:|  |: |:|:: ||  :|:||: |||         |
DSLWSAKVLVENPELIREVHLDYYRAGAQCAITASYQATPAGFAARGLDEAQSKALIGKSVELARKARE-----AYLAEN
              100       110       120       130       140       150

768       798       828       858       888       918       948       978
SERIYPLISGDVGPYAAFLADGSEYTGLYDIDKQGLKNFHRHRIELLLDEGVDILALETIPNAQEAEALIELLAEDFPQV
 :    |::| |||| |:||||||| | |    : :: ||| |:| ||| | |:|| ||:||  | ||| |||   :|:
PQAGTLLVAGSVGPYGAYLADGSEYRGDYHCSVEAFQAFHRPRVEALLDAGADLLACETLPNFSEIEALAELLT-AYPRA
         170       180       190       200       210       220       230

1008      1038      1068      1098      1128      1158      1188      1218
EAYMSFTSQDGKTISDGSAVADLAKAIDVSPQVVALGINCSSPSLVADFLQAIAEQTNKPLVTYPNSGEVYDGASQSWQS
 |: ||| :| : :|||: : |:   :   |||||||||| :       || :   |  ||| |||||| ||  |::|:
RAWFSFTLRDSEHLSDGTPLRDVVALLAGYPQVVALGINCIALENTTAALQHLHGLTVLPLVVYPNSGEHYDAVSKTWHH
          250       260       270       280       290       300       310

1248      1278      1308      1338      1368      1398      1428      1458
SPDHSHTLLENTSDWQKLGAQVVGGCCRTRPADIADLSAHLK*VKYLEEG*GKFDFLFQSTRKPAWILPNGFCFYLSEMT
  :|   | :    ||  ||:::|||||| ||||| | |
HGEHCAQLADYLPQWQAAGARLIGGCCRTTPADIAALKARS
          330       340       350
```

SEQ ID 8934 (GBS381) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 68 (lane 6; MW 42kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 4; MW 66.9kDa).

Example 1912

A DNA sequence (GBSx2021) was identified in *S.agalactiae* <SEQ ID 5927> which encodes the amino acid sequence <SEQ ID 5928>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2996(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1913

A DNA sequence (GBSx2022) was identified in *S.agalactiae* <SEQ ID 5929> which encodes the amino acid sequence <SEQ ID 5930>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.62   Transmembrane  176 - 192 ( 168 - 198)
   INTEGRAL    Likelihood =-11.57   Transmembrane   89 - 105 (  80 - 111)
   INTEGRAL    Likelihood =-10.03   Transmembrane  337 - 353 ( 332 - 359)
   INTEGRAL    Likelihood = -9.87   Transmembrane  292 - 308 ( 285 - 316)
   INTEGRAL    Likelihood = -4.51   Transmembrane   58 -  74 (  52 -  78)
   INTEGRAL    Likelihood = -3.88   Transmembrane  267 - 283 ( 267 - 286)
   INTEGRAL    Likelihood = -3.08   Transmembrane  125 - 141 ( 125 - 142)
   INTEGRAL    Likelihood = -2.13   Transmembrane  212 - 228 ( 212 - 228)

----- Final Results -----
              bacterial membrane --- Certainty=0.5649(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9377> which encodes amino acid sequence <SEQ ID 9378> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12034 GB:Z99105 similar to histidine permease [Bacillus subtilis]
 Identities = 221/384 (57%), Positives = 291/384 (75%), Gaps = 2/384 (0%)

Query: 2   PVTGSFHTYATKFISPGTGFTVAWLYWICWTVALGTEFLGAAMLMQRWFPNVPAWAFASF 61
            PVTG+FHTYA K+I PGTGFTVAWLYW+ WTVALG+EF  A +LMQRWFP+   W +++
Sbjct: 76  PVTGAFHTYAAKYIGPGTGFTVAWLYWLTWTVALGSEFTAAGLLMQRWFPHTSVWMWSAV 135

Query: 62  FALVIFGLNALSVRFFAEAESFFSSIKVIAIIIFIILGLGAMFGLVSFEGQHKAILFTHL 121
            FAL IF LNA SV+FFAE+E +FSSIKV+AI++FI+LG  AMFG++  +G   A + ++
Sbjct: 136 FALFIFLLNAFSVKFFAESEFWFSSIKVLAIVLFILLGGSAMFGIIPIKGGEAAPMLSNF 195

Query: 122 TANGA-FPNGIVAVVSVMLAVNYAFSGTELIGIAAGETDNPKEAVPRAIKTTIGRLVVFF 180
            TA G  FPNG V ++  ML+VN+AFSGTELIGIAAGE+ +P + +P+AIKTT+ RL +FF
Sbjct: 196 TAEGGLFPNGFVPILMTMLSVNFAFSGTELIGIAAGESVDPDKTIPKAIKTTVWRLSLFF 255

Query: 181 VLTIVVLASLLPMKEAGVSTAPFVDVFDKMGIPFTADIMNFVILTAILSAGNSGLYASSR 240
            V TI VL+ L+P+++AGV  +PFV VFD++G+P+ ADIMNFVILTAILSA NSGLYASSR
Sbjct: 256 VGTIFVLSGLIPIQDAGVIKSPFVAVFDRVGVPYAADIMNFVILTAILSAANSGLYASSR 315

Query: 241 MLWSLANEGMLSKSVVKINKHGVPMRALLLSMAGAVLSLFSSIYAADTVYLALVSIAGFA 300
            MLWSL+ E  L  +  K+   G P  AL+ SM G +LSL SS++A DTVY+ LVSI+GFA
Sbjct: 316 MLWSLSKEKTLHPTFAKLTSKGTPFNALVFSMIGGILSLLSSVFAPDTVYVVLVSISGFA 375

Query: 301 VVVVWLAIPVAQINFRKEFLKE-NQLEDLSYKTPFTPVLPYITIILLLISIVGIAWDSSQ 359
            VVVVW+ I  +Q  FRK +++  N++ DL Y+TP  P +P    +L L S+VGIA+D +Q
Sbjct: 376 VVVVWMGIAASQFMFRKRYIEAGNKVTDLKYRTPLYPFVPIAAFLLCLASVVGIAFDPNQ 435

Query: 360 RAGLYFGVPFIIFCYIYHKLRYKK 383
            R  LY GVPF+  CY  + ++ +K
Sbjct: 436 RIALYCGVPFMAICYAIYYVKNRK 459
```

There is also homology to SEQ ID 4070.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1914

A DNA sequence (GBSx2023) was identified in *S.agalactiae* <SEQ ID 5931> which encodes the amino acid sequence <SEQ ID 5932>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2378(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

There is also homology to SEQ ID 5642.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1915

A DNA sequence (GBSx2024) was identified in *S.agalactiae* <SEQ ID 5933> which encodes the amino acid sequence <SEQ ID 5934>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4935(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1916

A DNA sequence (GBSx2025) was identified in *S.agalactiae* <SEQ ID 5935> which encodes the amino acid sequence <SEQ ID 5936>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0530(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1917

A DNA sequence (GBSx2026) was identified in *S.agalactiae* <SEQ ID 5937> which encodes the amino acid sequence <SEQ ID 5938>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0175(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF63739 GB:AF236863 hypothetical GTP-binding protein
          [Lactococcus lactis]
 Identities = 142/193 (73%), Positives = 165/193 (84%)

Query: 6   LNTHNASILLSAANKSHYPQDDLPEVALAGRSNVGKSSFINTLLGRKNLARTSSKPGKTQ 65
            +NT+N +I +SAA+K  YP++D PE+ALAGRSNVGKSSFINTLL RKN ARTS +PGKTQ
Sbjct: 3   INTNNLTITISAASKKQYPENDWPEIALAGRSNVGKSSFINTLLNRKNFARTSGQPGKTQ 62

Query: 66  LLNFYNIDDKLRFVDVPGYGYAKVSKTERAKWGKMIEEYLVTRDNLRVVVSLVDFRHDPS 125
            LLNFYNIDD+L FVDVPGYGYA+VSK ER KWGKMIEEYL TR+NL+ VVSLVD RH+PS
Sbjct: 63  LLNFYNIDDQLHFVDVPGYGYARVSKKEREKWGKMIEEYLTTRENLKAVVSLVDIRHEPS 122

Query: 126 ADDIQMYEFLKYYEIPVIIVATKADKIPRGKWNKHESSIKKKLNFDKKDHFIVFSSVDRT 185
             DD+ MYEFLKYY IPVI+VATKADK+PRGKWNKHES IKK + FD  D FI+FSS D+T
Sbjct: 123 EDDLMMYEFLKYYHIPVILVATKADKVPRGKWNKHESIIKKAMKFDSTDDFIIFSSTDKT 182

Query: 186 GLDESWDTILSEL 198
            G++E+W  IL  L
Sbjct: 183 GIEEAWTAILKYL 195
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5939> which encodes the amino acid sequence <SEQ ID 5940>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0123(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 167/196 (85%), Positives = 183/196 (93%)

Query: 3   EEFLNTHNASILLSAANKSHYPQDDLPEVALAGRSNVGKSSFINTLLGRKNLARTSSKPG 62
            E+ LNTHNASILLSAANKSHYPQDDLPE+ALAGRSNVGKSSFINT+LGRKNLARTSSKPG
Sbjct: 4   EQVLNTHNASILLSAANKSHYPQDDLPEIALAGRSNVGKSSFINTILGRKNLARTSSKPG 63

Query: 63  KTQLLNFYNIDDKLRFVDVPGYGYAKVSKTERAKWGKMIEEYLVTRDNLRVVVSLVDFRH 122
            KTQLLNF+NIDDKLRFVDVPGYGYAKVSK+ERAKWGKMIEEYL +RDNLR VVSLVD RH
Sbjct: 64  KTQLLNFFNIDDKLRFVDVPGYGYAKVSKSERAKWGKMIEEYLTSRDNLRAVVSLVDLRH 123

Query: 123 DPSADDIQMYEFLKYYEIPVIIVATKADKIPRGKWNKHESSIKKKLNFDKKDHFIVFSSV 182
             PS +DIQMY+FLKYY+IPVI+VATKADKIPRGKWNKHES +KK LNFDK D FIVFSSV
Sbjct: 124 APSKEDIQMYDFLKYYDIPVIVVATKADKIPRGKWNKHESVVKKALNFDKSDTFIVFSSV 183
```

```
Query: 183 DRTGLDESWDTILSEL 198
           +R G+D+SWD IL ++
Sbjct: 184 ERIGIDDSWDAILEQV 199
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1918

A DNA sequence (GBSx2027) was identified in *S.agalactiae* <SEQ ID 5941> which encodes the amino acid sequence <SEQ ID 5942>. This protein is predicted to be protease ClpX (clpX). Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2389(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9947> which encodes amino acid sequence <SEQ ID 9948> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF63738 GB:AF236863 protease ClpX [Lactococcus lactis]
 Identities = 305/395 (77%), Positives = 357/395 (90%), Gaps = 1/395 (0%)

Query: 18  NVYCSFCGKSQDEVKKIIAGNGVFICNECVALSQEIIKEELAEEVLADLAEVPKPKELLE 77
            N+ CSFCGKSQD+VKK+IAG+ V+ICNEC+ LS  I++EEL EE  +++ EV  PKE+ +
Sbjct: 8   NIQCSFCGKSQDDVKKMIAGSDVYICNECIELSTRILEEELKEEQDSEMLEVKTPKEMFD 67

Query: 78  ILNQYVVGQDRAKRALAVAVYNHYKRVSYTESS-DDDVDLQKSNILMIGPTGSGKTFLAQ 136
             LN+YV+GQ++AKRALAVAVYNHYKR+++T S   +D++LQKSNIL+IGPTGSGKTFLAQ
Sbjct: 68  HLNEYVIGQEKAKRALAVAVYNHYKRINFTASKIAEDIELQKSNILLIGPTGSGKTFLAQ 127

Query: 137 TLAKSLNVPFAIADATSLTEAGYVGEDVENILLKLIQAADYNVERAERGIIYVDEIDKIA 196
            TLAKSLNVPFAIADATSLTEAGYVGEDVENILLKL+QA+D+N+ERAERGIIY+DEIDKIA
Sbjct: 128 TLAKSLNVPFAIADATSLTEAGYVGEDVENILLKLLQASDFNIERAERGIIYIDEIDKIA 187

Query: 197 KKGENVSITRDVSGEGVQQALLKIIEGTVASVPPQGGRKHPNQEMIQINTKNILFIVGGA 256
            KK ENVSITRDVSGEGVQQALLKIIEGTVASVPPQGGRKHPNQEMIQI+TKNILFIVGGA
Sbjct: 188 KKSENVSITRDVSGEGVQQALLKIIEGTVASVPPQGGRKHPNQEMIQIDTKNILFIVGGA 247

Query: 257 FDGIEDLVKQRLGEKVIGFGQTSRKIDDNASYMQEIISEDIQKFGLIPEFIGRLPVVAAL 316
            FDGIE++VKQRLGEK+IGFG  ++K+ D  SYMQEII+EDIQKFGLIPEFIGRLP+VAAL
Sbjct: 248 FDGIEEIVKQRLGEKIIGFGANNKKLSDEDSYMQEIIAEDIQKFGLIPEFIGRLPIVAAL 307

Query: 317 ELLTAEDLVRILTEPRNALVKQYQTLLSYDGVELEFDQDALLAIADKAIERKTGARGLRS 376
            E LT EDL++ILTEP+NAL+KQY+ LL +D VELEF   AL+AIA KAIERKTGARGLRS
Sbjct: 308 ERLTEEDLIQILTEPKNALIKQYKQLLLFDNVELEFKDGALMAIAKKAIERKTGARGLRS 367

Query: 377 IIEETMLDIMFEIPSQEDVTKVRITKAAVEGTDKP 411
            IIEE M+DIMFE+PS E++TKV IT+A V+G  +P
Sbjct: 368 IIEEVMMDIMFEVPSHEEITKVIITEAVVDGKAEP 402
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5943> which encodes the amino acid sequence <SEQ ID 5944>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
                   bacterial cytoplasm --- Certainty=0.2711(Affirmative) < succ>
                    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                     bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 378/409 (92%), Positives = 393/409 (95%), Gaps = 1/409 (0%)

Query: 9   MAGNRNNDMNVYCSFCGKSQDEVKKIIAGNGVFICNECVALSQEIIKEELAEEVLADLAE 68
           MAG+R ND+ VYCSFCGKSQD+VKKIIAGN VFICNECVALSQEIIKEELAEEVLADL E
Sbjct: 1   MAGSRTNDIKVYCSFCGKSQDDVKKIIAGNNVFICNECVALSQEIIKEELAEEVLADLTE 60

Query: 69  VPKPKELLEILNQYVVGQDRAKRALAVAVYNHYKRVSYTES-SDDDVDLQKSNILMIGPT 127
           VPKPKELL++LNQYVVGQDRAKRAL+VAVYNHYKRVS+TES  DDDVDLQKSNILMIGPT
Sbjct: 61  VPKPKELLDVLNQYVVGQDRAKRALSVAVYNHYKRVSFTESRDDDDVDLQKSNILMIGPT 120

Query: 128 GSGKTFLAQTLAKSLNVPFAIADATSLTEAGYVGEDVENILLKLIQAADYNVERAERGII 187
           GSGKTFLAQTLAKSLNVPFAIADATSLTEAGYVGEDVENILLKLIQAADYNVERAERGII
Sbjct: 121 GSGKTFLAQTLAKSLNVPFAIADATSLTEAGYVGEDVENILLKLIQAADYNVERAERGII 180

Query: 188 YVDEIDKIAKKGENVSITRDVSGEGVQQALLKIIEGTVASVPPQGGRKHPNQEMIQINTK 247
           YVDEIDKIAKKGENVSITRDVSGEGVQQALLKIIEGTVASVPPQGGRKHPNQEMIQI+TK
Sbjct: 181 YVDEIDKIAKKGENVSITRDVSGEGVQQALLKIIEGTVASVPPQGGRKHPNQEMIQIDTK 240

Query: 248 NILFIVGGAFDGIEDLVKQRLGEKVIGFGQTSRKIDDNASYMQEIISEDIQKFGLIPEFI 307
           NILFIVGGAFDGIE++VKQRLGEKVIGFGQ SRKIDDNASYMQEIISEDIQKFGLIPEFI
Sbjct: 241 NILFIVGGAFDGIEEIVKQRLGEKVIGFGQNSRKIDDNASYMQEIISEDIQKFGLIPEFI 300

Query: 308 GRLPVVAALELLTAEDLVRILTEPRNALVKQYQTLLSYDGVELEFDQDALLAIADKAIER 367
           GRLPVVAALE L   DL++ILTEPRNALVKQYQ LLSYDGVEL FD++AL AIA+KAIER
Sbjct: 301 GRLPVVAALEQLNTSDLIQILTEPRNALVKQYQALLSYDGVELAFDKEALEAIANKAIER 360

Query: 368 KTGARGLRSIIEETMLDIMFEIPSQEDVTKVRITKAAVEGTDKPVLETA 416
           KTGARGLRSIIEETMLDIMFEIPSQEDVTKVRITKAAVEG  KPVLETA
Sbjct: 361 KTGARGLRSIIEETMLDIMFEIPSQEDVTKVRITKAAVEGKSKPVLETA 409
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1919

A DNA sequence (GBSx2028) was identified in *S.agalactiae* <SEQ ID 5945> which encodes the amino acid sequence <SEQ ID 5946>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1920

A DNA sequence (GBSx2029) was identified in *S.agalactiae* <SEQ ID 5947> which encodes the amino acid sequence <SEQ ID 5948>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4029(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9949> which encodes amino acid sequence <SEQ ID 9950> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC33872 GB:AF055727 dihydrofolate reductase [Streptococcus pneumoniae]
 Identities = 83/162 (51%), Positives = 118/162 (72%), Gaps = 1/162 (0%)

Query: 25  MTKQIIAIWAEDEDHLIGVNGGLPWRLPKELHHFKETTMGQALLMGRKTFDGMNRRVLPG 84
            MTK+I+AIWA+DE+ LIG    LPW LP EL HFKETT+  A+LMGR TFDGM RR+LP
Sbjct: 1   MTKKIVAIWAQDEEGLIGKENRLPWHLPAELQHFKETTLNHAILMGRVTFDGMGRRLLPK 60

Query: 85  RETIILTKDEQFQADGVTVLNSVEQVIKWFQEHNKTLFIVGGASIYKAFLPYCEAIIKTK 144
            RET+ILT++ + + DGV     V+ V+ W+Q+  K L+I+GG  I++AF PY + +I T
Sbjct: 61  RETLILTRNPEEKIDGVATFQDVQSVLDWYQDQEKNLYIIGGKQIFQAFEPYLDEVIVTH 120

Query: 145 VHGKFKGDTYFP-DVNLSEFKVISRDYFEKDEQNAHAFTVTY 185
            +H + +GDTYFP +++LS F+ +S  ++ KDE+N + FT+ Y
Sbjct: 121 IHARVEGDTYFPEELDLSLFETVSSKFYAKDEKNPYDFTIQY 162
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5949> which encodes the amino acid sequence <SEQ ID 5950>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1214(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 82/160 (51%), Positives = 119/160 (74%)

Query: 25  MTKQIIAIWAEDEDHLIGVNGGLPWRLPKELHHFKETTMGQALLMGRKTFDGMNRRVLPG 84
            MTK+IIAIWAEDE  LIG+ G LPW LPKEL HFK+TT+ QA+LMGR TF+GMN + LP
Sbjct: 1   MTKEIIAIWAEDEAGLIGIAGKLPWYLPKELEHFKKTTLHQAILMGRVTFEGMNCKRLPQ 60

Query: 85  RETIILTKDEQFQADGVTVLNSVEQVIKWFQEHNKTLFIVGGASIYKAFLPYCEAIIKTK 144
            R+T+++T++  +Q D V  + S+E+V++W+   +KTL+I+GG  + +AF  Y + IIKT
Sbjct: 61  RQTLVMTRNRDYQVDEVLTMTSIEKVLEWYHAQDKTLYIIGGNKVLEAFNGYFDRIIKTV 120

Query: 145 VHGKFKGDTYFPDVNLSEFKVISRDYFEKDEQNAHAFTVT 184
            +H +FKGDTY P+++ S F   S+ ++ +D +N + FTVT
Sbjct: 121 IHHRFKGDTYRPNLDFSHFTQESQTFYARDAKNPYDFTVT 160
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1921

A DNA sequence (GBSx2030) was identified in *S.agalactiae* <SEQ ID 5951> which encodes the amino acid sequence <SEQ ID 5952>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1577(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA25221 GB:M33770 thymidylate synthase (EC 2.1.1.45)
          [Lactococcus lactis]
 Identities = 215/280 (76%), Positives = 245/280 (86%), Gaps = 2/280 (0%)

Query: 1   MTKADLLFKDNITKIMSEGVFSEQARPRYKNGEMANSKYITGAFAEYDLSKGEFPITTLR 60
           MT AD +FK NI  I+  GVFSE ARP+YK+G+MANSKY+TG+F  YDL KGEFPITTLR
Sbjct: 1   MTYADQVFKQNIQNILDNGVFSENARPKYKDGQMANSKYVTGSFVTYDLQKGEFPITTLR 60

Query: 61  PIPIKSAIKEIFWIYQDQTNDLAVLNDKYGVTYWNDWEVGHTGTIGQRYGAVVKKHNIIS 120
           PIPIKSAIKE+ WIYQDQT++L+VL +KYGV YW +W +G  GTIGQRYGA VKK+NII
Sbjct: 61  PIPIKSAIKELMWIYQDQTSELSVLEEKYGVKYWGEWGIGD-GTIGQRYGATVKKYNIIG 119

Query: 121 KLLKQLEDNPWNRRNVISLWDYEAFEETEGLLPCAFQTMFDVRRV-NGELYLDATLTQRS 179
           KLL+ L  NPWNRRN+I+LW YE FEETEGLLPCAFQTMFDVRR  +G++YLDATL QRS
Sbjct: 120 KLLEGLAKNPWNRRNIINLWQYEDFEETEGLLPCAFQTMFDVRREKDGQIYLDATLIQRS 179

Query: 180 NDMLVAHHINAMQYVALQMMIAKHFGWRVGKFFYFINNLHIYDNQFEQAQELLKRQPSEC 239
           NDMLVAHHINAMQYVALQMMIAKHF W+VGKFFYF+NNLHIYDNQFEQA EL+KR  SE
Sbjct: 180 NDMLVAHHINAMQYVALQMMIAKHFSWKVGKFFYFVNNLHIYDNQFEQANELMKRTASEK 239

Query: 240 NPKLVLNVPDGTDFFDIKPDDFALVDYDPIKPQLRFDLAI 279
            P+LVLNVPDGT+FFDIKP+DF LVDY+P+KPQL+FDLAI
Sbjct: 240 EPRLVLNVPDGTNFFDIKPEDFELVDYEPVKPQLKFDLAI 279
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5953> which encodes the amino acid sequence <SEQ ID 5954>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3131(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 227/279 (81%), Positives = 251/279 (89%)

Query: 1   MTKADLLFKDNITKIMSEGVFSEQARPRYKNGEMANSKYITGAFAEYDLSKGEFPITTLR 60
           MTKAD +FK NI KI++EG  SEQARP+YK+G  A+SKYITGAFAEYDL+KGEFPITTLR
Sbjct: 9   MTKADQIFKANIQKIINEGSLSEQARPKYKDGRTAHSKYITGAFAEYDLAKGEFPITTLR 68

Query: 61  PIPIKSAIKEIFWIYQDQTNDLAVLNDKYGVTYWNDWEVGHTGTIGQRYGAVVKKHNIIS 120
           PIPIKSAIKE+FWIYQDQ+N L VL  KY V YWN+WEV  T TIGQRYGAVVKKH+IIS
Sbjct: 69  PIPIKSAIKELFWIYQDQSNSLDVLEAKYNVHYWNEWEVDQTRTIGQRYGAVVKKHDIIS 128

Query: 121 KLLKQLEDNPWNRRNVISLWDYEAFEETEGLLPCAFQTMFDVRRVNGELYLDATLTQRSN 180
           K+LKQL +NPWNRRNVISLWDYEAFEET+GLLPCAFQ MFDVRRV  +LYLDA+LTQRSN
Sbjct: 129 KILKQLAENPWNRRNVISLWDYEAFEETKGLLPCAFQIMFDVRRVGEDLYLDASLTQRSN 188

Query: 181 DMLVAHHINAMQYVALQMMIAKHFGWRVGKFFYFINNLHIYDNQFEQAQELLKRQPSECN 240
           D+LVAHHINAMQYVALQMMIAKHFGW++GKFFYF+NNLHIYDNQF+QAQELLKRQP
Sbjct: 189 DILVAHHINAMQYVALQMMIAKHFGWKIGKFFYFVNNLHIYDNQFDQAQELLKRQPVASQ 248

Query: 241 PKLVLNVPDGTDFFDIKPDDFALVDYDPIKPQLRFDLAI 279
           PKLVLNVPD T+FFDIKPDDF L +YDP+KPQL FDLAI
Sbjct: 249 PKLVLNVPDRTNFFDIKPDDFELQNYDPVKPQLHFDLAI 287
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1922

A DNA sequence (GBSx2031) was identified in *S.agalactiae* <SEQ ID 5955> which encodes the amino acid sequence <SEQ ID 5956>. This protein is predicted to be HMG-CoA synthase. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0816(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5957> which encodes the amino acid sequence <SEQ ID 5958>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1670(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 260/385 (67%), Positives = 325/385 (83%)

Query: 36  MKIGIDKIGFATSQYVLEMTDLAIARQVDPEKFSKGLLLDSLSITPVTEDIVTLAASAAN 95
            M IGIDKIGFATSQYVL++ DLA+ARQVDP KFS+GLL++S S+ P+TEDI+TLAASAA+
Sbjct: 14  MTIGIDKIGFATSQYVLKLEDLALARQVDPAKFSQGLLIESFSVAPITEDIITLAASAAD 73

Query: 96  DILSDEDKETIDMVIVATESSIDQSKAASVYVHQLLEIQPFARSFEMKEACYSATAALDY 155
             IL+DED+  IDMVI+ATESS DQSKA+++YVH L+ IQPFARSFE+K+ACYSATAALDY
Sbjct: 74  QILTDEDRAKIDMVILATESSTDQSKASAIYVHHLVGIQPFARSFEVKQACYSATAALDY 133

Query: 156 AKLHVEKHPDSKVLVIASDIAKYGIKSTGESTQGAGSIAMLISQNPSILELKEDHLAQTR 215
            AKLHV   PDS+VLVIASDIA+YG+ S GESTQG+GSIA+L++ NP IL L ED++AQTR
Sbjct: 134 AKLHVASKPDSRVLVIASDIARYGVGSPGESTQGSGSIALLVTANPRILALNEDNVAQTR 193

Query: 216 DIMDFWRPNYSDVPYVNGMFSTKQYLDMLKTTWKVYQKRFNTSLSDYAAFCFHIPFPKLA 275
            DIMDFWRPNYS  PYV+G++STKQYL+ L+TTW+ YQKR N  LSD AA CFHIPFPKLA
Sbjct: 194 DIMDFWRPNYSFTPYVDGIYSTKQYLNCLETTWQAYQKRENLQLSDLAAVCFHIPFPKLA 253

Query: 276 LKGFNKILDNNLDEQKKAELQENFEHSITYSKKIGNCYTGSLYLGLLSLLENSQNLKAGD 335
            LKG N I+DN +  + + +L E F+ SI+YSK+IGN YTGSLYLGLLSLLENS+ L++GD
Sbjct: 254 LKGLNNIMDNTVPPEHREKLIEAFQASISYSKQIGNIYTGSLYLGLLSLLENSKVLQSGD 313

Query: 336 QIAFFSYGSGAVAEIFTGQLVDGYQNKLQSDRMDQLNKRQKITVTEYEKLFFEKTILDEN 395
            +I FFSYGSGAV+E ++GQLV GY   L ++R   L++R +++V++YE LF+E+  LD+N
Sbjct: 314 KIGFFSYGSGAVSEFYSGQLVAGYDKMLMTNRQALLDQRTRLSVSKYEDLFYEQVQLDDN 373

Query: 396 GNANFNTYRTGTFSLDSICEHQRIY 420
            GNANF+ Y TG F+L +I EH+RIY
Sbjct: 374 GNANFDIYLTGKFALTAIKEHRRIY 398
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1923

A DNA sequence (GBSx2032) was identified in *S.agalactiae* <SEQ ID 5959> which encodes the amino acid sequence <SEQ ID 5960>. This protein is predicted to be HMG-CoA reductase (mvaA). Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.49   Transmembrane  348 - 364 ( 348 - 364)
   INTEGRAL    Likelihood = -1.33   Transmembrane   53 -  69 (  53 -  69)

----- Final Results -----
               bacterial membrane --- Certainty=0.1595(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG02454 GB:AF290098 HMG-CoA reductase [Streptococcus pneumoniae]
 Identities = 266/421 (63%), Positives = 343/421 (81%), Gaps = 3/421 (0%)

Query: 3   KISWTGFSKKSPEERIHYLEEQDFLADSSLEIVTNQDLLSLSLANQMAENVIGRIALPFS 62
           KISW GFSKKS +ER+  L+ Q  L+      +   + +S+++A+Q++ENV+G  +LP+S
Sbjct: 2   KISWNGFSKKSYQERLELLKAQALLSPERQASLEKDEQMSVTVADQLSENVVGTFSLPYS 61

Query: 63  LVPDVLVNGKVYQVPYVTEEPSVVAAASFAAKIIKRSGGFLTTVHNRKMIGQVALYDVQD 122
           LVP+VLVNG+ Y VPYVTEEPSVVAAAS+A+KIIKR+GGF   VH R+MIGQVALY V +
Sbjct: 62  LVPEVLVNGQGYTVPYVTEEPSVVAAASYASKIIKRAGGFTAQVHQRQMIGQVALYQVAN 121

Query: 123 SQHTKESILNQKQQLLEIANAAHPSIVKRGGGACDLTIEI---KEDFLIVYLMVDTKEAM 179
            +  +E I ++K +LLE+AN A+PSIVKRGGGA DL +E    + DFL+VY+ VDT+EAM
Sbjct: 122 PKLAQEKIASKKAELLELANQAYPSIVKRGGGARDLHVEQIKGEPDFLVVYIHVDTQEAM 181

Query: 180 GANMVNTMMEALSSPLEDISKGKSLMSILSNYATESLVTATCRVDLRFLSRQKEEAIKLA 239
           GANM+NTM+EAL   LE++S+G+SLM ILSNYAT+SLVTA+CR+  R+LSRQK++  ++A
Sbjct: 182 GANMLNTMLEALKPVLEELSQGQSLMGILSNYATDSLVTASCRIAFRYLSRQKDQGREIA 241

Query: 240 QKMTMASQLAQVDPYRASTHNKGIFNGIDAIVLATGNDWRAIEAGAHTYAVKDGQYRGLS 299
           +K+ +ASQ AQ DPYRA+THNKGIFNGIDAI++ATGNDWRAIEAGAH +A +DG+Y+GLS
Sbjct: 242 EKIALASQFAQADPYRAATHNKGIFNGIDAILIATGNDWRAIEAGAHAFASRDGRYQGLS 301

Query: 300 RWSYKVDDNCLEGTLTLPMPVATKGGSIGINPSVHLAHDLLGRPNAKELASIILSIGLAQ 359
            W+  ++   L G +TLPMPVATKGGSIG+NP V L+HDLLG P+A+ELA II+SIGLAQ
Sbjct: 302 CWTLDLEREELVGEMTLPMPVATKGGSIGLNPRVALSHDLLGNPSARELAQIIVSIGLAQ 361

Query: 360 NFAALKALVSTGIQAGHMKLQAKSLALLAGAKEEQISEVVKQLLDSKHMNLETAQKIVNKL 420
           NFAALKALVSTGIQ GHMKLQAKSLALLAGA E +++  +V++L+  K  NLETAQ+ +  L
Sbjct: 362 NFAALKALVSTGIQQGHMKLQAKSLALLAGASESEVAPLVERLISDKTFNLETAQRYLENL 422
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5961> which encodes the amino acid sequence <SEQ ID 5962>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3929(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 257/422 (60%), Positives = 330/422 (77%)

Query: 2   TKISWTGFSKKSPEERIHYLEEQDFLADSSLEIVTNQDLLSLSLANQMAENVIGRIALPF 61
           T ++W+GFSKK+ EER+  +E+   L   +L  +    LL +  ANQM ENV+GR+ALPF
```

```
Sbjct: 4   TNLNWSGFSKKTFEERLQLIEKFKLLNAENLNQLKTDVLLPIQTANQMTENVLGRLALPF 63

Query: 62  SLVPDVLVNGKVYQVPYVTEEPSVVAAASFAAKIIKRSGGFLTTVHNRKMIGQVALYDVQ 121
           S+ PD LVNG  YQ+P+VTEEPSVVAAASFAAK+IKRSGGF     NR+MIGQ+ LYD+
Sbjct: 64  SIAPDFLVNGSTYQMPFVTEEPSVVAAASFAAKLIKRSGGFKAQTLNRQMIGQIVLYDID 123

Query: 122 DSQHTKESILNQKQQLLEIANAAHPSIVKRGGGACDLTIEIKEDFLIVYLMVDTKEAMGA 181
               + K +IL++ ++L+ +AN A+PSIVKRGGGA  + +E K +FLI YL VDT+EAMGA
Sbjct: 124 QIDNAKAAILHKTKKLIALANKAYPSIVKRGGGARTIHLEEKGEFLIFYLTVDTQEAMGA 183

Query: 182 NMVNTMMEALSSPLEDISKGKSLMSILSNYATESLVTATCRVDLRFLSRQKEEAIKLAQK 241
           NMVNTMMEAL   L  +SKG  LM+ILSNYATESLVT +C + +R L   K ++++LAQK
Sbjct: 184 NMVNTMMEALVPDLTRLSKGHCLMAILSNYATESLVTTSCEIPVRLLDHDKTKSLQLAQK 243

Query: 242 MTMASQLAQVDPYRASTHNKGIFNGIDAIVLATGNDWRAIEAGAHTYAVKDGQYRGLSRW 301
           + +AS+LAQVDPYRA+THNKGIFNGIDA+V+ATGNDWRAIEAGAH YA ++G Y+GLS+W
Sbjct: 244 IELASRLAQVDPYRATTHNKGIFNGIDAVVIATGNDWRAIEAGAHAYASRNGSYQGLSQW 303

Query: 302 SYKVDDNCLEGTLTLPMPVATKGGSIGINPSVHLAHDLLGRPNAKELASIILSIGLAQNF 361
            +  D   L G +TLPMP+A+KGGSIG+NP+V +AHDLL +P+AK LA +I S+GLAQNF
Sbjct: 304 HFDQDKQVLLGQMTLPMPIASKGGSIGLNPTVSIAHDLLNQPDAKTLAQLIASVGLAQNF 363

Query: 362 AALKALVSTGIQAGHMKLQAKSLALLAGAKEEQISEVVKQLLDSKHMNLETAQKIVNKLT 421
           AALKAL S+GIQAGHMKL AKSLALLAGA +++I+ +V  LL  K +NLE A   +++L
Sbjct: 364 AALKALTSSGIQAGHMKLHAKSLALLAGATQDEIAPLVNALLADKPINLEKAHFYLSQLR 423

Query: 422 KS 423
           +S
Sbjct: 424 QS 425
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1924

A DNA sequence (GBSx2033) was identified in *S.agalactiae* <SEQ ID 5963> which encodes the amino acid sequence <SEQ ID 5964>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2355(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5965> which encodes the amino acid sequence <SEQ ID 5966>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2687(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
  Identities = 76/138 (55%), Positives = 100/138 (72%), Gaps = 2/138 (1%)

Query: 7   PKWEELPELDLYLDQVLLYVNQLINPKTITNDKLLTASMINNYVKHNYISKPIKKKYNRR 66
           P W++LP+LDLYLDQVLLYVNQ  +   ++++K LTASMINNYVKH Y++KPIKKKY ++
```

```
Sbjct: 7   PYWKDLPDLDLYLDQVLLYVNQCTDFSEVSDNKSLTASMINNYVKHGYVTKPIKKKYQKQ 66

Query: 67  QVARLIVITAFKQVFAIQEISQTLELLTADNHSEEAYNGFAACMNKEE--VHDLPPVVIS 124
           Q+ARLI I+ FK VF IQ+IS+ LE L A   SE  YN F  C N++     D+PP+V
Sbjct: 67  QLARLIAISLFKTVFPIQDISRVLEELQAQADSESLYNTFVTCWNQKAPIEEDIPPIVQV 126

Query: 125 ACQTLNLYQETQKLVLEL 142
           ACQT+   Y +T  L+ E+
Sbjct: 127 ACQTVKDYHKTIYLLQEV 144
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1925

A DNA sequence (GBSx2034) was identified in *S.agalactiae* <SEQ ID 5967> which encodes the amino acid sequence <SEQ ID 5968>. This protein is predicted to be hemolysin iii. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.08   Transmembrane  142 - 158 ( 140 - 165)
   INTEGRAL    Likelihood = -6.79   Transmembrane   26 -  42 (  19 -  44)
   INTEGRAL    Likelihood = -5.63   Transmembrane  200 - 216 ( 196 - 217)
   INTEGRAL    Likelihood = -5.41   Transmembrane  104 - 120 ( 102 - 121)
   INTEGRAL    Likelihood = -3.98   Transmembrane   51 -  67 (  49 -  69)
   INTEGRAL    Likelihood = -1.86   Transmembrane  172 - 188 ( 169 - 188)

----- Final Results -----
               bacterial membrane --- Certainty=0.4630(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9951> which encodes amino acid sequence <SEQ ID 9952> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA58877 GB:X84058 novel hemolytic factor [Bacillus cereus]
 Identities = 79/204 (38%), Positives = 132/204 (63%), Gaps = 4/204 (1%)

Query: 17  EELANSITHAVGALLMLILLPITAVYSHNHFGLQAALGTSIFVTSLFLMFLSSSIYHSMT 76
           EE+AN+ITH +GA+L +  L I  +++  H    A +  +++  S+FL++L S++ HS+
Sbjct: 14  EEIANAITHGIGAILSIPALIILIIHASKHGTASAVVAFTVYGVSMFLLYLFSTLLHSIH 73

Query: 77  YNSLQKYVLRMIDHSMIYIAIAGSYTPVALSLIGGWLGYLIIFLQWGITLFGILYKIFAP 136
           +  ++K +  ++DHS IY+ IAG+YTP  L  + G LG+ ++ + W + + GI++KIF
Sbjct: 74  HPKVEK-LFTILDHSAIYLLIAGTYTPFLLITLRGPLGWTLLAIIWTLAIGGIIFKIFFV 132

Query: 137 KINDKFSLVLYLIMGWLVIF-IFPAIITKTGPAFWGLLLAGGICYTIGALFYA-RKRPYD 194
           +   K S + Y+IMGWL+I  I P     TG  F  LLLAGGI Y++GA+F+    K P++
Sbjct: 133 RRFIKASTLCYIIMGWLIIVAIKPLYENLTGHGF-SLLLAGGILYSVGAIFFLWEKLPFN 191

Query: 195 HMIWHLFILLASILQYIGIVYFML 218
           H IWHLF+L  S + +  +++++L
Sbjct: 192 HAIWHLFVLGGSAMMFFCVLFYVL 215
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5969> which encodes the amino acid sequence <SEQ ID 5970>. Analysis of this protein sequence reveals the following:

```
      Possible site: 41
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.51   Transmembrane  144 - 160 ( 138 - 163)
   INTEGRAL    Likelihood = -9.87   Transmembrane   49 -  65 (  45 -  71)
```

```
INTEGRAL    Likelihood = -7.11   Transmembrane  198 - 214  ( 193 - 215)
INTEGRAL    Likelihood = -6.16   Transmembrane  102 - 118  ( 100 - 120)
INTEGRAL    Likelihood = -2.97   Transmembrane   20 -  36  (  20 -  41)
INTEGRAL    Likelihood = -1.01   Transmembrane  167 - 183  ( 167 - 185)

----- Final Results -----
               bacterial membrane --- Certainty=0.5203(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA58877 GB:X84058 novel hemolytic factor [Bacillus cereus]
 Identities = 82/204 (40%), Positives = 128/204 (62%), Gaps = 4/204 (1%)

Query: 15  EEVANSVTHAIGAFAMLILLPISASYAYQTYDLKAAIGISIFVISLFLMFLSSTIYHSMA 74
            EE+AN++TH IGA   +  L I   +A +     A +  +++ +S+FL++L ST+ HS+
Sbjct: 14  EEIANAITHGIGAILSIPALIILIIHASKHGTASAVVAFTVYGVSMFLLYLFSTLLHSIH 73

Query: 75  YGSVHKYILRIIDHSMIYIAIAGSYTPVALSLVSGWLGYIIIVLQWGITLFGILYKIFAK 134
            +  V K +  I+DHS IY+ IAG+YTP  L  + G LG+ ++ + W + + GI++KIF
Sbjct: 74  HPKVEK-LFTILDHSAIYLLIAGTYTPFLLITLRGPLGWTLLAIIWTLAIGGIIFKIFFV 132

Query: 135 RINEKFSLMLYIVMGWL-VVFILPVIIQKTSLAFGLLMLFGGLSYTIGAVFYA-KKRPYF 192
            R   K S + YI+MGWL +V I P+    T   F LL L GG+ Y++GA+F+  +K P+
Sbjct: 133 RRFIKASTLCYIIMGWLIIVAIKPLYENLTGHGFSLL-LAGGILYSVGAIFFLWEKLPFN 191

Query: 193 HMIWHLFILLASALQFIAITFFML 216
            H IWHLF+L  SA+ F  + F++L
Sbjct: 192 HAIWHLFVLGGSAMMFFCVLFYVL 215
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 153/213 (71%), Positives = 181/213 (84%)

Query: 6   SIKLSPQLSFGEELANSITHAVGALLMLILLPITAVYSHNHFGLQAALGTSIFVTSLFLM 65
            + K S  LSF EE+ANS+THA+GA  MLILLPI+A Y++  + L+AA+G SIFV SLFLM
Sbjct: 4   TFKQSLPLSFSEEVANSVTHAIGAFAMLILLPISASYAYQTYDLKAAIGISIFVISLFLM 63

Query: 66  FLSSSIYHSMTYNSLQKYVLRMIDHSMIYIAIAGSYTPVALSLIGGWLGYLIIFLQWGIT 125
            FLSS+IYHSM Y S+ KY+LR+IDHSMIYIAIAGSYTPVALSL+ GWLGY+II LQWGIT
Sbjct: 64  FLSSTIYHSMAYGSVHKYILRIIDHSMIYIAIAGSYTPVALSLVSGWLGYIIIVLQWGIT 123

Query: 126 LFGILYKIFAPKINDKFSLVLYLIMGWLVIFIFPAIITKTGPAFWGLLLAGGICYTIGAL 185
            LFGILYKIFA +IN+KFSL+LY++MGWLV+FI P II KT  AF  L+L GG+ YTIGA+
Sbjct: 124 LFGILYKIFAKRINEKFSLMLYIVMGWLVVFILPVIIQKTSLAFGLLMLFGGLSYTIGAV 183

Query: 186 FYARKRPYDHMIWHLFILLASILQYIGIVYFML 218
            FYA+KRPY HMIWHLFILLAS LQ+I I +FML
Sbjct: 184 FYAKKRPYFHMIWHLFILLASALQFIAITFFML 216
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1926

A DNA sequence (GBSx2035) was identified in *S.agalactiae* <SEQ ID 5971> which encodes the amino acid sequence <SEQ ID 5972>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3641(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12492 GB:Z99107 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 81/302 (26%), Positives = 157/302 (51%), Gaps = 10/302 (3%)

Query: 1   MKSAYIFFNPKSGKDEQALAKEVKSYLIEHDFQDDY-VRIITPSSVEEAVALAKKASEDH 59
           MK A I +NP SG++   + K+  + +++   Q  Y       +   +A   AK+A+
Sbjct: 1   MKRARIIYNPTSGRE---IFKKHLAQVLQKFEQAGYETSTHATTCAGDATHAAKEAALRE 57

Query: 60  IDLVIPLGGDGTINKICGGVYAGGAYPTIGLVPAGTVNNFSKALNIPQERNL-ALENLLN 118
            DL+I  GGDGTIN++  G+      PT+G++P GT N+F++AL IP+E  L A + ++N
Sbjct: 58  FDLIIAAGGDGTINEVVNGLAPLDNRPTLGVIPVGTTNDFARALGIPREDILKAADTVIN 117

Query: 119 GHVKSVDICKVNDDYMISSLTLGLLADIAANVTSEMKRKLGPFAFLGDAYRILKRNRSYS 178
           G  + +DI +VN  Y I+    G L ++  +V S++K  LG  A+      +L   R
Sbjct: 118 GVARPIDIGQVNGQYFINIAGGGRLTELTYDVPSKLKTMLGQLAYYLKGMEMLPSLRPTE 177

Query: 179 ITLAYDNNVRSLRTRLLLITMTNSIAGMPAFSPEATIDDGLFRVYTMEHIHFFKLLLHLR 238
           + + YD  +      L L+T+TNS+ G    +P+++++DG+F +  ++  +  + +
Sbjct: 178 VEIEYDGKLFQGEIMLFLVTLTNSVGGFEKLAPDSSLNDGMFDLMILKKANLAEFIRVAT 237

Query: 239 QFRKGDFSQAKEIKHFHTNNLTISTFKRKKSAIPKVRIDGDPGDQLPVKVEVIPKALKFI 298
              +G+    + I +   N + ++  ++      ++ +DG+ G  LP +   + + +  +
Sbjct: 238 MALRGEHINDQHIIYTKANRVKVNVSEKM-----QLNLDGEYGGMLPGEFVNLYRHIHVV 292

Query: 299 IP 300
           +P
Sbjct: 293 MP 294
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5119> which encodes the amino acid sequence <SEQ ID 5120>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.4258(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 172/300 (57%), Positives = 229/300 (76%)

Query: 1   MKSAYIFFNPKSGKDEQALAKEVKSYLIEHDFQDDYVRIITPSSVEEAVALAKKASEDHI 60
           MK+  IF+NP SGK E  LA++VK Y  +H F +D V++ITP   ++A  LAK+A++D I
Sbjct: 1   MKTVRIFYNPNSGKKESQLARQVKDYFCQHGFSEDSVKVITPKDADQAFQLAKQAAKDKI 60

Query: 61  DLVIPLGGDGTINKICGGVYAGGAYPTIGLVPAGTVNNFSKALNIPQERNLALENLLNGH 120
           DLVIPLGGDGT+NKI GG+Y GGA+  IGLVP+GTVNNF+KA++IP +   AL+ +L G
Sbjct: 61  DLVIPLGGDGTLNKIIGGIYEGGAHCLIGLVPSGTVNNFAKAMHIPLQITEALDTILTGQ 120

Query: 121 VKSVDICKVNDDYMISSLTLGLLADIAANVTSEMKRKLGPFAFLGDAYRILKRNRSYSIT 180
           +K VDICK N  YMISSLTLGLLADIAA+VT+E KR+ GP AFL D+ RILKRNRSY+I+
Sbjct: 121 IKQVDICKANQQYMISSLTLGLLADIAADVTAEEKRRFGPLAFLKDSIRILKRNRSYAIS 180

Query: 181 LAYDNNVRSLRTRLLLITMTNSIAGMPAFSPEATIDDGLFRVYTMEHIHFFKLLLHLRQF 240
           L   N+   L+T+ LLITMTN+IAG P+FSP A  DDG F+VYTM+ + FFK L H+  F
Sbjct: 181 LISHNHRIHLKTKFLLITMTNTIAGFPSFSPGAQADDGYFQVYTMKKVSFFKFLWHINDF 240

Query: 241 RKGDFSQAKEIKHFHTNNLTISTFKRKKSAIPKVRIDGDPGDQLPVKVEVIPKALKFIIP 300
           ++GDFS+A+EI HF  N L++    +K++ +P+ RIDGD  D LP+++++IPKA+  I+P
Sbjct: 241 KQGDFSKAEEISHFQANTLSLLPQAKKQAILPRTRIDGDKSDYLPIQLDIIPKAVSIIVP 300
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1927

A DNA sequence (GBSx2036) was identified in *S.agalactiae* <SEQ ID 5973> which encodes the amino acid sequence <SEQ ID 5974>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3628(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB10885 GB:AB010693 gene_id:K21C13.21~pir||T04769~strong
           similarity to unknown protein [Arabidopsis thaliana]
 Identities = 85/291 (29%), Positives = 150/291 (51%), Gaps = 28/291 (9%)

Query: 10  DQEWEVPVESGRYHMIVGEFCPYAQRPQIARQLLGLDKHISISFVDDV------------ 57
            D + + P ESGRYH+ +   CP+A R     ++ GLD+ I+ S V  +
Sbjct: 29  DPDSQFPAESGRYHLYISYACPWACRCLSYLKIKGLDEAITFSSVHAIWGRTKETDDHRG 88

Query: 58  ----PSDIGLIFSQPEQVTGAKSLRDIYHLTDPTYQGPYTIPILIDKTDNRIVCKESADL 113
                 SD  L  ++P+ + GAKS+R++Y +  P Y+G YT+P+L DK    +V  ES+++
Sbjct: 89  WVFPDSDTELPGAEPDYLNGAKSVRELYEIASPNYEGKYTVPVLWDKKLKTVVNNESSEI 148

Query: 114 LRLFTTDFSDLHQEDAPVLFSQETASLIDNDIKDINKNFQSLMYKLAFLDKQADYDTYSK 173
            +R+F T+F+ + +  +  L+      +I+     +     + +YK  F  KQ  Y+
Sbjct: 149 IRMFNTEFNGIAKTPSLDLYPSHLRDVINETNGWVFNGINNGVYKCGFARKQEPYNEAVN 208

Query: 174 EFFTFLDQKEHLLGQRPFLLGDNLSEVDIHFFTPLVRWDIAGRDLLLLNQKALEDYPNIF 233
            + +  +D+ E +LG++ ++ G+  +E DI  F  L+R+D         N++ L +YPNIF
Sbjct: 209 QLYEAVDRCEEVLGKQRYICGNTFTEADIRLFVTLIRFDEVYAVHFKCNKRLLREYPNIF 268

Query: 234 SWAKTLYNDFNLKTLTNPQSIKNNYY-----LGKFGRAVRHHTIVPTGPNM 279
            ++ K +Y    + +  N + IK +YY     +  FG       I+P GPN+
Sbjct: 269 NYIKDIYQIHGMSSTVNMEHIKQHYYGSHPTINPFG-------IIPHGPNI 312
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1928

A DNA sequence (GBSx2037) was identified in *S.agalactiae* <SEQ ID 5975> which encodes the amino acid sequence <SEQ ID 5976>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2647(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07793 GB:AB037666 hypothetical protein [Streptomyces sp.
           CL190]
```

```
 Identities = 127/331 (38%), Positives = 194/331 (58%), Gaps = 9/331 (2%)

Query: 4   RKDDHIKYALKYQSHY---NSFDDIELIHSSLPKYNVNDIDLSTHFAGQSFEFPFYINAM 60
            RKDDH++ A++  + +   N FDD+  +H +L   +  D+ L+T FAG S++ P YINAM
Sbjct: 6   RKDDHVRLAIEQHNAHSGRNQFDDVSFVHHALAGIDRPDVSLATSFAGISWQVPIYINAM 65

Query: 61  TGGSEKGKAVNHKLAQVAQATGIVMVTGSYSAALKNDE--DDSYPTTDLYPDLKLATNIG 118
            TGGSEK   +N  LA  A+ TG+ + +GS +A +K+    D      D  P+  +  NI
Sbjct: 66  TGGSEKTGLINRDLATAARETGVPIASGSMNAYIKDPSCADTFRVLRDENPNGFVIANIN 125

Query: 119 LDKPVPAAESTVKAMNPIFLQVHVNVMQELLMPEGEREFHMWRSHLKEYVDNIQCPLILK 178
                V  A+  +  +    LQ+H+N  QE  MPEG+R F  W   +++    +  P+I+K
Sbjct: 126 ATTTVDNAQRAIDLIEANALQIHINTAQETPMPEGDRSFASWVPQIEKIAAAVDIPVIVK 185

Query: 179 EVGFGMDLQSIKDAYDIGITTVDISGRGGTSFAYIENQRGR--DRSYLNTWGQTTAQSLI 236
            EVG G+  Q+I    D+G+   D+SGRGGT FA IEN R    D ++L+ WGQ+TA  L+
Sbjct: 186 EVGNGLSRQTILLLADLGVQAADVSGRGGTDFARIENGRRELGDYAFLHGWGQSTAACLL 245

Query: 237 NAQSMMDKMDILASGGIRHPLDMVKCLVLGAKAVGLSRTVLELVERYPVDDVIAILNSWK 296
            +AQ +   + +LASGG+RHPLD+V+ L LGA+AVG S   L  +    VD +I  L +W
Sbjct: 246 DAQDI--SLPVLASGGVRHPLDVVRALALGARAVGSSAGFLRTLMDDGVDALITKLTTWL 303

Query: 297 EDLRMIMCALNCKKITDLRQVNYILYGQLKE 327
            + L  +   L  +   DL + + +L+G+L++
Sbjct: 304 DQLAALQTMLGARTPADLTRCDVLLHGELRD 334
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5977> which encodes the amino acid sequence <SEQ ID 5978>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2823(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 244/329 (74%), Positives = 284/329 (86%)

Query: 1   MTNRKDDHIKYALKYQSHYNSFDDIELIHSSLPKYNVNDIDLSTHFAGQSFEFPFYINAM 60
            MTNRKDDHIKYALKYQS YN+FDDIELIH SLP Y+++DIDLSTHFAGQ F+FPFYINAM
Sbjct: 31  MTNRKDDHIKYALKYQSPYNAFDDIELIHHSLPSYDLSDIDLSTHFAGQDFDFPFYINAM 90

Query: 61  TGGSEKGKAVNHKLAQVAQATGIVMVTGSYSAALKNDEDDSYPTTDLYPDLKLATNIGLD 120
            TGGS+KGKAVN KLA+VA ATGIVMVTGSYSAALKN  DDSY   ++  +LKLATNIGLD
Sbjct: 91  TGGSQKGKAVNEKLAKVAAATGIVMVTGSYSAALKNPNDDSYRLHEVADNLKLATNIGLD 150

Query: 121 KPVPAAESTVKAMNPIFLQVHVNVMQELLMPEGEREFHMWRSHLKEYVDNIQCPLILKEV 180
            KPV    + TV+ M P+FLQVHVNVMQELLMPEGER FH W+ HL EY   I  P+ILKEV
Sbjct: 151 KPVALGQQTVQEMQPLFLQVHVNVMQELLMPEGERVFHTWKKHLAEYASQIPVPVILKEV 210

Query: 181 GFGMDLQSIKDAYDIGITTVDISGRGGTSFAYIENQRGRDRSYLNTWGQTTAQSLINAQS 240
            GFGMD+ SIK A+D+GI T DISGRGGTSFAYIENQRG DRSYLN WGQTT Q L+NAQ
Sbjct: 211 GFGMDVNSIKLAHDLGIQTFDISGRGGTSFAYIENQRGGDRSYLNDWGQTTVQCLLNAQG 270

Query: 241 MMDKMDILASGGIRHPLDMVKCLVLGAKAVGLSRTVLELVERYPVDDVIAILNSWKEDLR 300
            +MD+++ILASGG+RHPLDM+KC VLGA+AVGLSRTVLELVE+YP + VIAI+N WKE+L+
Sbjct: 271 LMDQVEILASGGVRHPLDMIKCFVLGARAVGLSRTVLELVEKYPTERVIAIVNGWKEELK 330

Query: 301 MIMCALNCKKITDLRQVNYILYGQLKEAN 329
            +IMCAL+CK I +L+ V+Y+LYG+L++ N
Sbjct: 331 IIMCALDCKTIKELKGVDYLLYGRLQQVN 359
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1929

A DNA sequence (GBSx2038) was identified in *S.agalactiae* <SEQ ID 5979> which encodes the amino acid sequence <SEQ ID 5980>. This protein is predicted to be phosphomevalonate kinase. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0785(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG02457 GB:AF290099 phosphomevalonate kinase [Streptococcus pneumoniae]
 Identities = 170/330 (51%), Positives = 233/330 (70%), Gaps = 1/330 (0%)

Query: 1   MVKVQTGGKLYIAGEYAILYPGQVAILKNVPIYMTALATFADNYSLYSDMFNYTASLQPD 60
           M+ V+T GKLY AGEYAIL PGQ+A++K++PIYM A   F+D+Y +YSDMF++   L+P+
Sbjct: 1   MIAVKTCGKLYWAGEYAILEPGQLALIKDIPIYMRAEIAFSDSYRIYSDMFDFAVDLRPN 60

Query: 61  KQYSLIQETILLMEEWLINFGKNIKPIHLEITGKLERYGLKFGIGSSGSVVVLTIKAMAA 120
             YSLIQETI LM ++L   G+N++P  L+I GK+ER G KFG+GSSGSVVVL +KA+ A
Sbjct: 61  PDYSLIQETIALMGDFLAVRGQNLRPFSLKICGKMEREGKKFGLGSSGSVVVLVVKALLA 120

Query: 121 LYEIEMPSDLLFKLSAYVLLKRGDNGSMGDIACIAYEHLISYSAFDRRAVSKMIETKPLE 180
           LY + +  +LLFKL++ VLLKRGDNGSMGD+ACI  E L+ Y +FDR+  +  +E + L
Sbjct: 121 LYNLSVDQNLLFKLTSAVLLKRGDNGSMGDLACIVAEDLVLYQSFDRQKAAAWLEEENLA 180

Query: 181 QVLEAEWGYRITKIQALLEMDFLVGWTMQPSISKEMINIVKSTITQRFLDDTKYQVVQLL 240
            VLE +WG+ I++++  LE DFLVGWT + ++S  M+  +K  I Q FL  +K  VV L+
Sbjct: 181 TVLERDWGFFISQVKPTLECDFLVGWTKEVAVSSHMVQQIKQNINQNFLSSSKETVVSLV 240

Query: 241 SAFKEGDKEAIKRCLEEISLLLFNLHPSIYTDKLQKLKEASKGLDIVTKSSGSGGGDCGI 300
            A ++G  E +   +E  S LL  L    IYT  L++LKEAS+ L  V KSSG+GGGDCGI
Sbjct: 241 EALEQGKAEKVIEQVEVASKLLEGLSTDIYTPLLRQLKEASQDLQAVAKSSGAGGGDCGI 300

Query: 301 AISFN-KNDNQTLIKRWESAGIELLSKETL 329
           A+SF+ ++   TL  RW   GIELL +E +
Sbjct: 301 ALSFDAQSSRNTLKNRWADLGIELLYQERI 330
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5981> which encodes the amino acid sequence <SEQ ID 5982>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2669(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
  Identities = 171/325 (52%), Positives = 227/325 (69%), Gaps = 2/325 (0%)

Query: 4   VQTGGKLYIAGEYAILYPGQVAILKNVPIYMTALATFADNYSLYSDMFNYTASLQPDKQY 63
           VQTGGKLY+ GEYAIL PGQ A++  +P+ MTA +  A +  L SDMF++ A + PD  Y
Sbjct: 22  VQTGGKLYLTGEYAILTPGQKALIHFIPLMMTAEISPAAHIQLASDMFSHKAGMTPDASY 81
```

```
Query: 64  SLIQETILLMEEWLINFGKNIKPIHLEITGKLERYGLKFGIGSSGSVVVLTIKAMAALYE 123
           +LIQ T+    ++L      ++P  L ITGK+ER G KFGIGSSGSV +LT+KA++A Y+
Sbjct: 82  ALIQATVKTFADYLGQSIDQLEPFSLIITGKMERDGKKFGIGSSGSVTLLTLKALSAYYQ 141

Query: 124 IEMPSDLLFKLSAYVLLKRGDNGSMGDIACIAYEHLISYSAFDRRAVSKMIETKPLEQVL 183
           I +  +LLFKL+AY LLK+GDNGSMGDIACIAY+ L++Y++FDR  VS  ++T PL+++L
Sbjct: 142 ITLTPELLFKLAAYTLLKQGDNGSMGDIACIAYQTLVAYTSFDREQVSNWLQTMPLKKLL 201

Query: 184 EAEWGYRITKIQALLEMDFLVGWTMQPSISKEMINIVKSTITQRFLDDTKYQVVQ-LLSA 242
             +WGY I  IQ  L  DFLVGWT  P+IS++MI  V ++IT  FL  T YQ+ Q  + A
Sbjct: 202 VKDWGYHIQVIQPALPCDFLVGWTKIPAISRQMIQQVTASITPAFL-RTSYQLTQSAMVA 260

Query: 243 FKEGDKEAIKRCLEEISLLLFNLHPSIYTDKLQKLKEASKGLDIVTKSSGSGGGDCGIAI 302
            +EG KE +K+ L   S LL  LHP+IY  KL  L  A +  D V KSSGSGGGDCGIA+
Sbjct: 261 LQEGHKEELKKSLAGASHLLKELHPAIYHPKLVTLVAACQKQDAVAKSSGSGGGDCGIAL 320

Query: 303 SFNKNDNQTLIKRWESAGIELLSKE 327
           +FN++   TLI +W+ A I LL +E
Sbjct: 321 AFNQDARDTLISKWQEADIALLYQE 345
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1930

A DNA sequence (GBSx2039) was identified in *S.agalactiae* <SEQ ID 5983> which encodes the amino acid sequence <SEQ ID 5984>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.75    Transmembrane    20 -  36 (  18 -  36)

----- Final Results -----
               bacterial membrane --- Certainty=0.1702(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1931

A DNA sequence (GBSx2040) was identified in *S.agalactiae* <SEQ ID 5985> which encodes the amino acid sequence <SEQ ID 5986>. This protein is predicted to be mevalonate diphosphate decarboxylase. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1557(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG02456 GB:AF290099 mevalonate diphosphate decarboxylase
           [Streptococcus pneumoniae]
```

```
 Identities = 219/312 (70%); Positives = 264/312 (84%)

Query: 1   MDGKSISVKSYANIAIIKYWGKADAEKMIPATSSISLTLENMYTETRLTALGKDAKKDEF 60
           MD + ++V+SYANIAIIKYWGK   ++M+PATSSISLTLENMYTET L+ L  +   DEF
Sbjct: 1   MDREPVTVRSYANIAIIKYWGKKKEKEMVPATSSISLTLENMYTETTLSPLPANVTADEF 60

Query: 61  YISGVLQNDHEHDKMSAILDRFRQNRSGFVKIETTNNMPTAAGLSSSSSGLSALVKACND 120
           YI+G LQN+ EH KMS I+DR+R    GFV+I+T NNMPTAAGLSSSSSGLSALVKACN
Sbjct: 61  YINGQLQNEVEHAKMSKIIDRYRPAGEGFVRIDTQNNMPTAAGLSSSSSGLSALVKACNA 120

Query: 121 FFGTNLSQSQLAQEAKFASGSSSRSFFGPVAAWDKDSGDIYKVHTNLDLAMIMLVLNDKR 180
           +F   L +SQLAQEAKFASGSSSRSF+GP+ AWDKDSG+IY V T+L LAMIMLVL DK+
Sbjct: 121 YFKLGLDRSQLAQEAKFASGSSSRSFYGPLGAWDKDSGEIYPVETDLKLAMIMLVLEDKK 180

Query: 181 KPISSREGMKICTETSTTFNEWVRQSEQDYQDMLVYLKNNDFQKVGQLTERNALAMHSTT 240
           KPISSR+GMK+C ETSTTF++WVRQSE+DYQDML+YLK NDF K+G+LTE+NALAMH+TT
Sbjct: 181 KPISSRDGMKLCVETSTTFDDWVRQSEKDYQDMLIYLKENDFAKIGELTEKNALAMHATT 240

Query: 241 KTATPAFSYLTEETYKAMDVVKKLREKGHECYYTMDAGPNVKVLCLRQDLEALAAILEKD 300
           KTA+PAFSYLT+ +Y+AM  V++LREKG  CY+TMDAGPNVKV C  +DLE L+ I  +
Sbjct: 241 KTASPAFSYLTDASYEAMAFVRQLREKGEACYFTMDAGPNVKVFCQEKDLEHLSEIFGQR 300

Query: 301 YRIIVSTTKELA 312
           YR+IVS TK+L+
Sbjct: 301 YRLIVSKTKDLS 312
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5987> which encodes the amino acid sequence <SEQ ID 5988>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1271(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 221/313 (70%), Positives = 258/313 (81%)

Query: 1   MDGKSISVKSYANIAIIKYWGKADAEKMIPATSSISLTLENMYTETRLTALGKDAKKDEF 60
           +D   I+V SYANIAIIKYWGK +  KMIP+TSSISLTLENM+T T ++ L   A  D+F
Sbjct: 1   VDPNVITVTSYANIAIIKYWGKENQAKMIPSTSSISLTLENMFTTTSVSFLPDTATSDQF 60

Query: 61  YISGVLQNDHEHDKMSAILDRFRQNRSGFVKIETTNNMPTAAGLSSSSSGLSALVKACND 120
           YI+G+LQND EH K+SAI+D+FRQ    FVK+ET NNMPTAAGLSSSSSGLSALVKAC+
Sbjct: 61  YINGILQNDEEHTKISAIIDQFRQPGQAFVKMETQNNMPTAAGLSSSSSGLSALVKACDQ 120

Query: 121 FFGTNLSQSQLAQEAKFASGSSSRSFFGPVAAWDKDSGDIYKVHTNLDLAMIMLVLNDKR 180
            F T L Q  LAQ+AKFASGSSSRSFFGPVAAWDKDSG IYKV T+L +AMIMLVLN  +
Sbjct: 121 LFDTQLDQKALAQKAKFASGSSSRSFFGPVAAWDKDSGAIYKVETDLKMAMIMLVLNAAK 180

Query: 181 KPISSREGMKICTETSTTFNEWVRQSEQDYQDMLVYLKNNDFQKVGQLTERNALAMHSTT 240
           KPISSREGMK+C +TSTTF++WV QS  DYQ ML YLK N+F+KVGQLTE NALAMH+TT
Sbjct: 181 KPISSREGMKLCRDTSTTFDQWVEQSAIDYQHMLTYLKTNNFEKVGQLTEANALAMHATT 240

Query: 241 KTATPAFSYLTEETYKAMDVVKKLREKGHECYYTMDAGPNVKVLCLRQDLEALAAILEKD 300
           KTA P FSYLT+E+Y+AM+ VK+LR++G  CY+TMDAGPNVKVLCL +DL  LA  L K+
Sbjct: 241 KTANPPFSYLTKESYQAMEAVKELRQEGFACYFTMDAGPNVKVLCLEKDLAQLAERLGKN 300

Query: 301 YRIIVSTTKELAD 313
           YRIIVS TK+L D
Sbjct: 301 YRIIVSKTKDLPD 313
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1932

A DNA sequence (GBSx2041) was identified in *S.agalactiae* <SEQ ID 5989> which encodes the amino acid sequence <SEQ ID 5990>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1512(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5991> which encodes the amino acid sequence <SEQ ID 5992>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1117(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 182/290 (62%), Positives = 223/290 (76%)

Query: 1   MKEKFGIGKAHSKIILMGEHSVVYGYPAIAIPLKNIEVTCLIEEAPQLIALDMTDPLSTA 60
           M E  G GKAHSKIIL+GEH+VVYGYPAIA+PL +IEV C I  A + +  D  D LSTA
Sbjct: 6   MNENIGYGKAHSKIILIGEHAVVYGYPAIALPLTDIEVVCHIFPADKPLVFDFYDTLSTA 65

Query: 61  IFAALDYLGKTSSKIAYHIESQVPERRGMGSSAAVAIAAIRAVFDYFDEDLEADLLECLV 120
           I+A+LDYL +    IAY I SQVP++RGMGSSAAV+IAAIRAVF Y  E L  DLLE LV
Sbjct: 66  IYASLDYLQRLQEPIAYEIVSQVPQKRGMGSSAAVSIAAIRAVFSYCQEPLSDDLLEILV 125

Query: 121 NRAEMIAHSNPSGLDAKTCLSENTIKFIRNIGFSTVPMHLNAYLVIADTGIHGHTKEAVD 180
           N+AE+IAH+NPSGLDAKTCLS++ IKFIRNIGF T+ + LN YL+IADTGIHGHT+EAV+
Sbjct: 126 NKAEIIAHTNPSGLDAKTCLSDHAIKFIRNIGFETIEIALNGYLIIADTGIHGHTREAVN 185

Query: 181 KVKSSGEAVLPFLKELGYLAEASEDAIHKSDSKQLGSLMTKAHQSLKQLGVSSLEADHLV 240
           KV    E  LP+L +LG L +A E AI++ +   +G LMT+AH +LK +GVS  +AD LV
Sbjct: 186 KVAQFEETNLPYLAKLGALTQALERAINQKNKVAIGQLMTQAHSALKAIGVSISKADQLV 245

Query: 241 EVAISCGALGAKMSGGGLGGCIIALVKEKREAERLSQQLEREGAVNTWTE 290
           E A+  GALGAKM+GGGLGGC+IAL   K  AE++S +L+ EGAVNTW +
Sbjct: 246 EAALRAGALGAKMTGGGLGGCMIALADTKDMAEKISHRLKEEGAVNTWIQ 295
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1933

A DNA sequence (GBSx2042) was identified in *S.agalactiae* <SEQ ID 5993> which encodes the amino acid sequence <SEQ ID 5994>. This protein is predicted to be a histidine protein kinase. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-13.43    Transmembrane    12 -  28 (   4 -  33)
```

```
   INTEGRAL    Likelihood = -9.29   Transmembrane  163 - 179 ( 157 - 191)

----- Final Results -----
               bacterial membrane --- Certainty=0.6371(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF79919 GB:AF039082 putative histidine protein kinase
           [Lactococcus lactis]
 Identities = 78/315 (24%), Positives = 154/315 (48%), Gaps = 33/315 (10%)

Query: 101 SDRQIKNYAKRIVSQNSHSGHITYNFSTYSYLLKKVGKNDYLVVFLDTTNQYLDNQRLLQ 160
           +++QI N  + +  +N +  +  Y + T S  +          V++   +     Q  +
Sbjct: 84  NEKQI-NTIQTVSVKNPYGDNWHYRYLTTSQFIITNSDGTVTPVYVQIFSNVDQIQDAMS 142

Query: 161 LSIWM---SLVSFIVFMVIVSV-LSGRVILPFVANYEKQRRFITNAGHELKTPLAIISAN 216
            ++W+   ++++F +  VI+S+ L+   + P +A YEKQ+ F+ NA HEL+TPLAI+
Sbjct: 143 RAMWVIVTTMITFWILSVIISLYLANWTLKPILAAYEKQKEFVENASHELRTPLAILQNR 202

Query: 217 NELV-----EMMSGESEWTKSTNDQIQRLTGLINGMVSLAR------FEEQPDISM---- 261
            EL+       +  +SE    +  +++ +  L + +++LAR       E +P  +
Sbjct: 203 LELLFQKPTATIIDQSENISESLSEVRNMRLLTSNLLNLARRDSGIKIEPEPTTATYFEN 262

Query: 262 VDLDFSHITKDAAEDFKGPIIKDGKDFIMSIQPGIHVKAEEKSLFELVTLLVDNANKYCD 321
           +   +  +T++A + F G +  +G            V  ++  + +L+T+L DNA KY D
Sbjct: 263 IFNSYEMLTENAGKKFSGNLKLEGT-----------VNLDQALIKQLLTILFDNALKYTD 311

Query: 322 PMGTVTVKLSRSSRLRRAKLEVSNTYKNGKDIDYSKFFERFYREDESHNNKKSGYGIGLS 381
             G ++V + ++         V++  +   D D  K F+RF+R D++   +K G G+GLS
Sbjct: 312 SEGEISVDVIKNGGF--LTFAVADNGEGISDEDKKKIFDRFFRVDKARTRQKGGLGLGLS 369

Query: 382 IVTSLVHLFKGSIDV 396
           +   +V  + G I V
Sbjct: 370 LAKQIVEAYNGKITV 384
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5751> which encodes the amino acid sequence <SEQ ID 5752>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.30   Transmembrane   18 -  34 (  13 -  42)
   INTEGRAL    Likelihood =-10.35   Transmembrane  170 - 186 ( 163 - 199)

----- Final Results -----
               bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 233/410 (56%), Positives = 303/410 (73%), Gaps = 1/410 (0%)

Query: 1   MFRNLRLRFIGIAALAILVVLFSVVGVLNSANHYQTKNEIYRVLTILADNNGRIPNKLEF 60
           MF  +R+RFI IA++AI ++L S+VG++N+A  YQ++ EI R+L +++ N G++P   E
Sbjct: 10  MFNRIRIRFIMIASIAIFIILSSIVGIINTARCYQSQQEINRILHLISSNKGKLPGTTES 69

Query: 61  SKELGDDLSTDAIFQFRYFSARTDAKGNVTSFDSRNIFEVSDRQIKNYAKRIVSQNSHSG 120
           SK LG  LS D++ QFRY+S   +A G++ S ++ NI  +   + + +A+         G
Sbjct: 70  SKRLGTKLSEDSLSQFRYYSVIFNANGHLLSSNTANISALDREEAQYFARLFAKSGEEKG 129

Query: 121 HITYNFSTYSYLLKKVGKNDYLVVFLDTTNQYLDNQRLLQLSIWMSLVSFIVFMVIVSVL 180
              +  S YSYL+ ++   + LVV LDTT  +     LL +S+ ++   FI F+V+VS+
Sbjct: 130 SYRHQDSVYSYLITQLPNEEKLVVILDTTFYFRSVGDLLAVSVMLAFGGFIFFVVLVSLF 189

Query: 181 SGRVILPFVANYEKQRRFITNAGHELKTPLAIISANNELVEMMSGESEWTKSTNDQIQRL 240
```

```
           SG VI PFV NYEKQRRFITNAGHELKTPLAIISANNELVE+M+GESEWTKST+DQ++RL
Sbjct: 190 SGMVIKPFVQNYEKQRRFITNAGHELKTPLAIISANNELVELMTGESEWTKSTSDQVKRL 249

Query: 241 TGLINGMVSLARFEEQPDISMVDLDFSHITKDAAEDFKGPIIKDGKDFIMSIQPGIHVKA 300
           TGLIN M++LAR EEQPD+ +  +DFS I +DAAEDFK  ++KDGK F ++IQP I +KA
Sbjct: 250 TGLINQMITLARLEEQPDVVLHMVDFSAIAQDAAEDFKSLVLKDGKRFDLTIQPNIMIKA 309

Query: 301 EEKSLFELVTLLVDNANKYCDPMGTVTVKLSRSSRLR-RAKLEVSNTYKNGKDIDYSKFF 359
           EEKSLFELVT+LVDNANKYCDP G V V L+   R R RAKLEVSNTY  GK IDYS+FF
Sbjct: 310 EEKSLFELVTILVDNANKYCDPKGLVKVSLTTIGRRRKRAKLEVSNTYLEGKSIDYSRFF 369

Query: 360 ERFYREDESHNNKKSGYGIGLSIVTSLVHLFKGSIDVNYKHDTITFVIYI 409
           ERFYREDESHN+K+ GYGIGLS+  S+V LFKG+I VNYK+D I F + I
Sbjct: 370 ERFYREDESHNSKEKGYGIGLSMAESMVKLFKGTITVNYKNDAIVFTVVI 419
```

SEQ ID 5994 (GBS273) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 14; MW 46kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 56 (lane 5; MW 71kDa).

GBS273-GST was purified as shown in Figure 208, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1934

A DNA sequence (GBSx2043) was identified in *S.agalactiae* <SEQ ID 5995> which encodes the amino acid sequence <SEQ ID 5996>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2181(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1935

A DNA sequence (GBSx2044) was identified in *S.agalactiae* <SEQ ID 5997> which encodes the amino acid sequence <SEQ ID 5998>. This protein is predicted to be two-component response regulator (trcR). Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2503(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9379> which encodes amino acid sequence <SEQ ID 9380> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04091 GB:AP001508 two-component response regulator [Bacillus halodurans]
 Identities = 71/183 (38%), Positives = 120/183 (64%), Gaps = 3/183 (1%)

Query: 9    RVLIAEDEEQMSRVLSTAISHQGYVVDVAYDGQTAIDLANQNAYDVMVMDVMMPVKTGIE 68
             R+LI EDE++++RVL    + H+GY  D A+ G    ++     +A+D++++DVM+P  +G+E
Sbjct: 3    RILIIEDEKKIARVLQLELEHEGYETDAAFSGSDGLETFQAHAWDLVLLDVMLPELSGLE 62

Query: 69   AVKEIRQSGNKSHIIMLTAMAEIDDRVTGLDAGADDYLTKPFSLKELLARLRSMSRRLE- 127
              ++ IR +    + II+LTA   I D+V+GLD GA+DY+TKPF ++ELLAR+R+  R ++
Sbjct: 63   VLRRIRMTDPVTPIILLTARNSIPDKVSGLDLGANDYITKPFEIEELLARVRACLRTVQT 122

Query: 128  -DFTPNVLSLGRVTLSVGEQELQCEN-TIRLAGKEAKMLAFFMLNHDKELSTQQLFEHVW 185
             +    + L    +T++    +++Q  N TI L  KE ++L FF+ N  + LS +Q+  +VW
Sbjct: 123  RERVEDTLMFQELTINEKTRDVQRGNETIELTPKEFELLVFFIKNKGQVLSREQILTNVW 182

Query: 186  GAD 188
             G D
Sbjct: 183  GFD 185
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 5999> which encodes the amino acid sequence <SEQ ID 6000>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2391(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 125/185 (67%), Positives = 151/185 (81%)

Query: 8    MRVLIAEDEEQMSRVLSTAISHQGYVVDVAYDGQTAIDLANQNAYDVMVMDVMMPVKTGI 67
             M++L+AEDE QMS VL+TA++HQGY VDV ++GQ AID A  NAYD+M++D+MMP+K+GI
Sbjct: 1    MKILLAEDEWQMSNVLTTAMTHQGYDVDVVFNGQEAIDKAKDNAYDIMILDIMMPIKSGI 60

Query: 68   EAVKEIRQSGNKSHIIMLTAMAEIDDRVTGLDAGADDYLTKPFSLKELLARLRSMSRRLE 127
             EA+KEIR SGN SHIIMLTAMAEI+DRVTGLDAGADDYLTKPFSLKELLARLRSM RR+E
Sbjct: 61   EALKEIRASGNCSHIIMLTAMAEINDRVTGLDAGADDYLTKPFSLKELLARLRSMERRVE 120

Query: 128  DFTPNVLSLGRVTLSVGEQELQCENTIRLAGKEAKMLAFFMLNHDKELSTQQLFEHVWGA 187
              FTP VL    VTL++ EQEL   N IRLA KE K++AF MLN  K L T+ L++HVW
```

```
                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB05604 GB:AP001513 unknown conserved protein [Bacillus halodurans]
 Identities = 67/182 (36%), Positives = 111/182 (60%), Gaps = 4/182 (2%)

Query: 17  LEDFSQRIQLENDKAKVETGYKLYEHIIGRIKTSDSMIEKCRRKQLPVTVDSALKTIRDS 76
            L++ + +I +   + +    Y   EH+  R+K+ +S++ K +R+    T++S  + +RD
Sbjct: 29  LQELNTKIDILKQEFQYIHDYNPIEHVSSRVKSPESIVNKIQRRGNDFTLESIRENVRDI 88

Query: 77  IGVRIICGFVNDIYQIIERIKAFDDCRIVVEKDYIQHVKPNGYRSYHVILEIDTPYPDCL 136
             G+RI C F +DIY + E++    D  +V  KDYI++ KPNGYRS H+IL I    P  +
Sbjct: 89  AGIRITCSFESDIYTLSEQLMQQHDISVVETKDYIKNPKPNGYRSLHLILSI----PIFM 144

Query: 137 GNSDGKYYIEIQLRTIAQDSWASLEHQMKYKHDIENPERIVRELKRCADEMASVDLTMQT 196
             +     Y+E+Q+RTIA D WASLEH++ YK++   PE +++ELK  A+  A +D  M+
Sbjct: 145 SDRVQDVYVEVQIRTIAMDFWASLEHKIYYKYNKNVPEHLLKELKDAAESAALLDQKMEK 204

Query: 197 IR 198
            I+
Sbjct: 205 IQ 206
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6003> which encodes the amino acid sequence <SEQ ID 6004>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1057(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 127/206 (61%), Positives = 162/206 (77%)

Query: 3   TNIYGDYGRYLPLILEDFSQRIQLENDKAKVETGYKLYEHIIGRIKTSDSMIEKCRRKQL 62
            ++IY  +  YLPL+L+  +  I  EN K+K ETG+KLYEH   RIK+  SMIEKC+RKQL
Sbjct: 11  SSIYSGFEVYLPLVLQTITDVIIAENIKSKKETGFKLYEHFTSRIKSEASMIEKCQRKQL 70

Query: 63  PVTVDSALKTIRDSIGVRIICGFVNDIYQIIERIKAFDDCRIVVEKDYIQHVKPNGYRSY 122
            P+T  SALK I+DSIG+RIICGF++DIY++++ +K+     +  EKDYI + KPNGYRSY
Sbjct: 71  PLTSKSALKIIKDSIGIRIICGFIDDIYRMVDLLKSIPGMSVNTEKDYILNAKPNGYRSY 130

Query: 123 HVILEIDTPYPDCLGNSDGKYYIEIQLRTIAQDSWASLEHQMKYKHDIENPERIVRELKR 182
            H+ILE++T +PD LG   G Y+IE+QLRTIAQDSWASLEHQMKYKH + N E I RELKR
Sbjct: 131 HLILELETHFPDILGEKKGCYFIEVQLRTIAQDSWASLEHQMKYKHQVANAEMITRELKR 190

Query: 183 CADEMASVDLTMQTIRQLIESGTKKE 208
            CADE+AS D+TMQTIRQLI+  T++E
Sbjct: 191 CADELASCDVTMQTIRQLIQETTEEE 216
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1937

A DNA sequence (GBSx2046) was identified in *S.agalactiae* <SEQ ID 6005> which encodes the amino acid sequence <SEQ ID 6006>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3250(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA37193 GB:X53013 ORF1 (AA 1 - 384) [Lactococcus lactis]
 Identities = 30/55 (54%), Positives = 37/55 (66%)

Query: 1   MEFYYKTLKRKFINDADTIFIEQSQFEIFIYIETDHNSSSSHVVLDYQSQKEFEK 55
           ME +YKTLKR+ INDA      ++  EIF YIET +N+   H  LDYQS K+FEK
Sbjct: 327 MESFYKTLKRELINDAHFETRAEATQEIFKYIETYYNTKWMHSGLDYQSPKDFEK 381
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6007> which encodes the amino acid sequence <SEQ ID 6008>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3065(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 31/59 (52%), Positives = 39/59 (65%)

Query: 1  MEFYYKTLKRKFINDADTIFIEQSQFEIFIYIETDHNSSSSHVVLDYQSQKEFEKIITN 59
          ME +YKTLKR+ +NDA    I+Q+Q EIF Y ET +N    H  L Y S  EFEKI+T+
Sbjct: 13 MEAFYKTLKRELVNDAHFATIKQAQLEIFKYSETYYNPKRLHSALGYLSPVEFEKIVTH 71
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1938

A DNA sequence (GBSx2047) was identified in *S.agalactiae* <SEQ ID 6009> which encodes the amino acid sequence <SEQ ID 6010>. This protein is predicted to be R5 protein. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.98    Transmembrane   30 -  46 (  29 -  51)
   INTEGRAL    Likelihood = -2.76    Transmembrane  967 - 983 ( 966 - 985)

----- Final Results -----
                bacterial membrane --- Certainty=0.2593(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8935> which encodes amino acid sequence <SEQ ID 8936> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1    Crend: 8
SRCFLG: 0
McG: Length of UR:    2
     Peak Value of UR:   2.44
     Net Charge of CR: 2
McG: Discrim Score:      0.78
GvH: Signal Score (-7.5): -0.0599995
```

```
     Possible site: 39
>>> Seems to have a cleavable N-term signal seq.
Amino Acid Composition: calculated from 40
ALOM program   count: 0 value:   7.37 threshold:  0.0
   PERIPHERAL  Likelihood =  7.37     194
 modified ALOM score:  -1.97

*** Reasoning Step: 3

Rule gpol

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif: 944-948
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 8936 (GBS200) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 3; MW 107.4kDa), in Figure 169 (lane 4; MW 122kDa) and in Figure 238 (lane 11; MW 122kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 35 (lane 3; MW 132kDa).

Purified Thio-GBS200-His is shown in Figure 244, lane 9.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1939

A DNA sequence (GBSx2048) was identified in *S.agalactiae* <SEQ ID 6011> which encodes the amino acid sequence <SEQ ID 6012>. This protein is predicted to be a 16.1 kDa transcriptional regulator. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3919(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9953> which encodes amino acid sequence <SEQ ID 9954> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB16108 GB:Z99124 similar to transcriptional regulator (MarR
           family) [Bacillus subtilis]
 Identities = 30/114 (26%), Positives = 59/114 (51%), Gaps = 3/114 (2%)

Query: 29  DVEHLAGPQGHLVMYLYKHPDKDMSIKAVEEILHISKSVASNLVKRMEKNGFIAIVPSKT 88
            D++   G   +LV  +Y++P   +  + + E++ + ++ A+  +K++E  GFI  +P +
Sbjct: 25  DLDLTRGQYLYLVR-IYENPG--IIQEKLAEMIKVDRTTAARAIKKLEMQGFIQKLPDEQ 81

Query: 89  DKRVKYLYLTHLGKKKATQFEIFLEKLHSTMLAGITKEEIRTTKKVIRTLAKNM 142
            +K++K L+ T  GKK         E      L+G T EE  T   ++  + KN+
Sbjct: 82  NKKIKKLFPTEKGKKVYPLLRREGEHSTEVALSGFTSEEKETISALLHRVRKNI 135
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6013> which encodes the amino acid sequence <SEQ ID 6014>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4175(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 27/64 (42%), Positives = 46/64 (71%)

Query: 3  MENPLQKARILVNQLEKYLDHYAKEYDVEHLAGPQGHLVMYLYKHPDKDMSIKAVEEILH 62
           M   +   R L++Q+E+  D  AK+YDVEHLAGPQG+++++L KH ++++ +K +E+ L
Sbjct: 1  MSQVIGDLRELIHQIEQISDEIAKKYDVEHLAGPQGYVLVFLAKHQNQEIFVKDIEKQLR 60

Query: 63 ISKS 66
           I +S
Sbjct: 61 IFQS 64
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1940

A DNA sequence (GBSx2049) was identified in *S.agalactiae* <SEQ ID 6015> which encodes the amino acid sequence <SEQ ID 6016>. This protein is predicted to be 5'-nucleotidase family protein. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -2.66    Transmembrane   668 - 684 ( 665 - 684)

----- Final Results -----
               bacterial membrane --- Certainty=0.2062(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12747 GB:Z99108 similar to 5'-nucleotidase [Bacillus subtilis]
 Identities = 178/535 (33%), Positives = 270/535 (50%), Gaps = 55/535 (10%)

Query: 28   DQVGVQVIGVNDFHGALDNTGTANMPDGKVANAGTAAQLD---AYMDDAQKDFKQTNPNG 84
            + V ++++ +ND HG +D     ++ DG     GT  ++D   AY+ + + + K
Sbjct: 586  EHVPLRILSMNDLHGKIDQQYELDL-DGNGTVDGTFGRMDYAAAYLKEKKAEKKN----- 639

Query: 85   ESIRVQAGDMVGASPANSGLLQDEPTVKNFNAMNVEYGTLGNHEFDEGLAEYNRIVTGKA 144
             S+ V AGDM+G S   S LLQDEPTV+    +  + GT+GNHEFDEG  E  RI+ G
Sbjct: 640  -SLIVHAGDMIGGSSPVSSLLQDEPTVELMEDIGFDVGTVGNHEFDEGTDELLRILNG-G 697

Query: 145  PAPDSNINNITKSYPHEAAKQEIVVANVIDKVNKQIPYNWKPYAIKNIPVNNKSVNVGFI 204
              P        +++P       +V AN       ++    +P+      +N + V V FI
Sbjct: 698  DHPKGTSGYDGQNFP-------LVCANC------KMKSTGEPFLPAYDIINVEGVPVAFI 744

Query: 205  GIVTKDIPNLVLRKNYEQYEFLDEAETIVKYAKELQAKNVKAIVVLAHVPATSKNDIAEG 264
            G+VT+    +V+ +  +  EF DEA  + K A+EL+ K VKAI VLAH+ A    +   G
Sbjct: 745  GVVTQSAAGMVMPEGIKNIEFTDEATAVNKAAEELKKKGVKAIAVLAHMSAEQNGNAITG 804

Query: 265  EAAEMMKKVNQLFPENSVDIVFAGHNHQYTNGLVGKTRIVQALSQGKAYADVRGVLDTDT 324
            E+A++  K      ++ +D++FA HNHQ  NG V    IVQA   GKA   V   +D  T
```

```
Sbjct: 805  ESADLANKT-----DSEIDVIFAAHNHQVVNGEVNGKLIVQAFEYGKAIGVVDVEIDKTT 859

Query: 325  QDFIETPSAKVIAVAPGKKTGSADIQAIVDQANTIVKQVTEAKIGTAEVSVMITRSVDQD 384
            +D ++  SA+++ V   K        AI+ +  TI + +    +G A V +    S D D
Sbjct: 860  KDIVK-KSAEIVYVDQSKIEPDVSASAILKKYETIAEPIISEVVGEAAVDMEGGYSNDGD 918

Query: 385  NVSPVGSLITEAQLAIARKSWPDIDFAMTNNGGIRADLLIKPDGTITWGAAQAVQPFGNI 444
              +P+G+LI +   A  +      DFA+ N GGIR  L     G ITWG    +QPFGN+
Sbjct: 919  --TPLGNLIADGMRAAMK-----TDFALMNGGGIREAL---KKGPITWGDLYNIQPFGNV 968

Query: 445  LQVVEITGRDLYKALNEQYDQKQNFFLQIAGLRYTYTDNKEGGEETPFKVVKAYKSNGEE 504
            L  +EI G+DL + +N Q          I+G  +TYT +KE G+    K+     ++G E
Sbjct: 969  LTKLEIKGKDLREIINAQISPVFGPDYSISG--FTYTWDKETGKAVDMKM-----ADGTE 1021

Query: 505  INPDAKYKLVINDFLFGGGDGFASFRNAKLLGAINP-----DTEVFMAYITDLEK 554
            I PDA Y L +N+F+       A ++   LLG  NP     D E  + Y+   ++
Sbjct: 1022 IQPDATYTLTVNNFMATATG--AKYQPIGLLGK-NPVTGPEDLEATVEYVKSFDE 1073
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1607> which encodes the amino acid sequence <SEQ ID 1608>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.67   Transmembrane  662 - 678 ( 661 - 679)
   INTEGRAL    Likelihood = -2.02   Transmembrane   19 -  35 (  18 -  35)

----- Final Results -----
               bacterial membrane --- Certainty=0.2869(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 415/688 (60%), Positives = 517/688 (74%), Gaps = 21/688 (3%)

Query: 1    MKKKIILKSSVLGLVAGTSIMFSSVFADQVGVQVIGVNDFHGALDNTGTANMPDGKVANA 60
            MKK  ILKSSVL ++   +++ + V ADQV VQ +GVNDFHGALDNTGTA  P GK+ NA
Sbjct: 14   MKKYFILKSSVLSILTSFTLLVTDVQADQVDVQFLGVNDFHGALDNTGTAYTPSGKIPNA 73

Query: 61   GTAAQLDAYMDDAQKDFKQTNPNGESIRVQAGDMVGASPANSGLLQDEPTVKNFNAMNVE 120
            GTAAQL AYMDDA+ DFKQ N +G SIRVQAGDMVGASPANS LLQDEPTVK FN M  E
Sbjct: 74   GTAAQLGAYMDDAEIDFKQANQDGTSIRVQAGDMVGASPANSALLQDEPTVKVFNKMKFE 133

Query: 121  YGTLGNHEFDEGLAEYNRIVTGKAPAPDSNINNITKSYPHEAAKQEIVVANVIDKVNKQI 180
            YGTLGNHEFDEGL E+NRI+TG+AP P+S IN+ITK Y HEA+ Q IV+ANVIDK  K I
Sbjct: 134  YGTLGNHEFDEGLDEFNRIMTGQAPDPESTINDITKQYEHEASHQTIVIANVIDKKTKDI 193

Query: 181  PYNWKPYAIKNIPVNNKSVNVGFIGIVTKDIPNLVLRKNYEQYEFLDEAETIVKYAKELQ 240
            PY WKPYAIK+I +N+K V +GFIG+VT +IPNLVL++NYE Y+FLD AETI KYAKELQ
Sbjct: 194  PYGWKPYAIKDIAINDKIVKIGFIGVVTTEIPNLVLKQNYEHYQFLDVAETIAKYAKELQ 253

Query: 241  AKNVKAIVVLAHVPATSKNDIAEGEAAEMMKKVNQLFPENSVDIVFAGHNHQYTNGLVGK 300
             ++V AIVVLAHVPATSK+ + + E A +M+KVNQ++PE+S+DI+FAGHNHQYTNG +GK
Sbjct: 254  EQHVHAIVVLAHVPATSKDGVVDHEMATVMEKVNQIYPEHSIDIIFAGHNHQYTNGTIGK 313

Query: 301  TRIVQALSQGKAYADVRGVLDTDTQDFIETPSAKVIAVAPGKKTGSADIQAIVDQANTIV 360
            TRIVQALSQGKAYADVRG LDTDT DFI+TPSA V+AVAPG KT ++DI+AI++ AN IV
Sbjct: 314  TRIVQALSQGKAYADVRGTLDTDTNDFIKTPSANVVAVAPGIKTENSDIKAIINHANDIV 373

Query: 361  KQVTEAKIGTAEVSVMITRSVDQDNVSPVGSLITEAQLAIARKSWPDIDFAMTNNGGIRA 420
            K VTE KIGTA  S  I+++ + D  SPVG+L T AQL IA+K++P +DFAMTNNGGIR+
Sbjct: 374  KTVTERKIGTATNSSTISKTENIDKESPVGNLATTAQLTIAKKTFPTVDFAMTNNGGIRS 433

Query: 421  DLLIKPDGTITWGAAQAVQPFGNILQVVEITGRDLYKALNEQYDQKQNFFLQIAGLRYTY 480
            DL++K D TITWGAAQAVQPFGNILQV+++TG+ +Y  LN+QYD+ Q +FLQ++GL YTY
Sbjct: 434  DLVVKNDRTITWGAAQAVQPFGNILQVIQMTGQHIYDVLNQQYDENQTYFLQMSGLTYTY 493
```

```
Query: 481 TDNKEGGEETPFKVVKAYKSNGEEINPDAKYKLVINDFLFGGGDGFASFRNAKLLGAINP 540
           TDN     +TPFK+VK YK NGEEIN    Y +V+NDFL+GGGDGF++F+ AKL+GAIN
Sbjct: 494 TDNDPKNSDTPFKIVKVYKDNGEEINLTTTYTVVVNDFLYGGGDGFSAFKKAKLIGAINT 553

Query: 541 DTEVFMAYITDLEKAGKKVSVPNNKPKIYVTMKMVNETITQNDGTHSIIKKLYLDRQGNI 600
           DTE F+ YIT+LE +GK V+      K YVT  + + T   + G HSII K++ +R GN
Sbjct: 554 DTEAFITYITNLEASGKTVNATIKGVKNYVTSNLESSTKVNSAGKHSIISKVFRNRDGNT 613

Query: 601 VAQEIVSDTLNQTKSKSTKINPVTTIHKKQLHQFTAINPMRNYGKPSNSTTVKSKQLPKT 660
           V+ E++SD L  T++ +  +    T                     +N  T+ S  LP T
Sbjct: 614 VSSEVISDLLTSTENTNNSLGKKET--------------------TTNKNTISSSTLPIT 653

Query: 661 NSEYGQSFLMSVFG-VGLIGIALNTKKK 687
              Y  S +M++   + L G+    KK+
Sbjct: 654 GDNYKMSPIMTILALISLGGLNAFIKKR 681
```

SEQ ID 6016 (GBS328) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 69 (lane 4; MW 73kDa). The GBS328-His fusion product was purified (Figure 213, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 268), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1941

A DNA sequence (GBSx2050) was identified in *S.agalactiae* <SEQ ID 6017> which encodes the amino acid sequence <SEQ ID 6018>. This protein is predicted to be peptide deformylase (def-2). Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.70   Transmembrane   55 -  71 (  55 -  74)

----- Final Results -----
               bacterial membrane --- Certainty=0.1680(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB09662 GB:Z96934 peptide deformylase [Clostridium
           beijerinckii]
 Identities = 71/136 (52%), Positives = 96/136 (70%)

Query: 1   MIKPIVRDTFFLQQKSQMASRADVSLAKDLQETLHANQNYCVGMAANMIGSLKRVIIINV 60
           MIKPIV+D  FL QKS+ A++ D+ +  DL +TL AN  +CVG+AANMIG  KR+++  V
Sbjct: 1   MIKPIVKDILFLGQKSEEATKNDMVVIDDLIDTLRANLEHCVGLAANMIGVKKRILVFTV 60

Query: 61  GITNLVMFNPVVVAKSDPYETEESCLSLVGCRSTQRYCHITISYRDINWKEQQIKLTDFP 120
           G   + M NPV++ K  PYETEESCLSL+G R T+RY  I ++Y D N+ +++     F
Sbjct: 61  GNLIVPMINPVILKKEKPYETEESCLSLIGFRKTKRYETIEVTYLDRNFNKKKQVFNGFT 120

Query: 121 AQICQHELDHLEGILI 136
           AQI QHE+DH EGI+I
Sbjct: 121 AQIIQHEMDHFEGIII 136
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6019> which encodes the amino acid sequence <SEQ ID 6020>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence
```

```
INTEGRAL    Likelihood = -3.61    Transmembrane    55 -   71 (  55 -  73)

----- Final Results -----
               bacterial membrane --- Certainty=0.2444(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 77/136 (56%), Positives = 103/136 (75%)

Query: 1   MIKPIVRDTFFLQQKSQMASRADVSLAKDLQETLHANQNYCVGMAANMIGSLKRVIIINV 60
           MI+ I+ D F LQQK+Q+A + D+ + +DLQ+TL   +  C+GMAANMIG  KR++I+++
Sbjct: 1   MIREIITDHFLLQQKAQVAKKEDLWIGQDLQDTLAFYRQECLGMAANMIGEQKRIVIVSM 60

Query: 61  GITNLVMFNPVVVAKSDPYETEESCLSLVGCRSTQRYCHITISYRDINWKEQQIKLTDFP 120
           G  +LVMFNPV+V+K   Y+T+ESCLSL G R TQRY  IT+ Y D NW+ +++ LT
Sbjct: 61  GFIDLVMFNPVMVSKKGIYQTKESCLSLSGYRKTQRYDKITVEYLDHNWRPKRLSLTGLT 120

Query: 121 AQICQHELDHLEGILI 136
           AQICQHELDHLEGILI
Sbjct: 121 AQICQHELDHLEGILI 136
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1942

A DNA sequence (GBSx2051) was identified in *S.agalactiae* <SEQ ID 6021> which encodes the amino acid sequence <SEQ ID 6022>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2880(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB05820 GB:AP001514 NADP-specific glutamate dehydrogenase
          [Bacillus halodurans]
 Identities = 298/444 (67%), Positives = 362/444 (81%), Gaps = 2/444 (0%)

Query: 7   YVASVLEKVKKQNEHEEEFLQAVEEVFESLVPVFDKYPQYIEENLLERLVEPERVISFRV 66
           YV  V E VK++N +E EF QAV+EVF+SL+PV  K+PQY+++ +LER+VEPERVISFRV
Sbjct: 16  YVQHVYETVKRRNPNEHEFHQAVKEVFDSLLPVLVKHPQYVKQAILERIVEPERVISFRV 75

Query: 67  PWVDDKGQVQVNRGYRVQFSSAIGPYKGGLRFHPTVTQSIVKFLGFEQIFKNSLTGLPIG 126
           PWVDD+G VQVNRG+RVQF+SA+GPYKGGLRFHP+V  SI+KFLGFEQIFKN+LTG PIG
Sbjct: 76  PWVDDQGNVQVNRGFRVQFNSALGPYKGGLRFHPSVNASIIKFLGFEQIFKNALTGQPIG 135

Query: 127 GGKGGSNFDPKGKSDNEVMRFTQSFMTELQKYIGPDLDVPAGDIGVGGREIGYLYGQYKR 186
           GGKGGS+FDPKGKSD E+MRF+QSFM+EL  YIGPD+DVPAGDIGVG +EIGY++GQYK+
Sbjct: 136 GGKGGSDFDPKGKSDGEIMRFSQSFMSELSNYIGPDIDVPAGDIGVGAKEIGYMFGQYKK 195

Query: 187 L-NGYQNGVLTGKGLTYGGSLARTEATGYGAVYFAKEMLAARGQDLTGKVALVSGSGNVA 245
           +  G++ GVLTGKG+ YGGSLAR EATGYG VYF +EM+   G    G   +VSGSGNV+
Sbjct: 196 MRGGFEAGVLTGKGIGYGGSLARKEATGYGTVYFVEEMIKDHGFSFAGSTVVVSGSGNVS 255

Query: 246 IYATEKLQELGATVVAVSDSSGYVYDPDGIDLETLKQIKEVERARIVKYTEKHPKANFTP 305
           IYA EK  +LGA VVA SDS GYVYD +GIDL+T+K++KEVER RI +Y  +HP A++
Sbjct: 256 IYAMEKAMQLGAKVVACSDSGGYVYDKNGIDLQTVKRLKEVERKRISEYVNEHPHAHYVQ 315

Query: 306 ADQGSIWSIKADLAFPCATQNELDEEDAKLLVENGVLAVTEGANMPSTLGAIKVFQKAGV 365
```

```
                  G IWS+  D+A PCATQNELDE  A +L+ NGV AV EGANMPSTL A+  FQ+ GV
Sbjct: 316 GCSG-IWSVPCDIALPCATQNELDEAAATMLIANGVKAVGEGANMPSTLQAVHTFQEHGV 374

Query: 366 AFGPAKAANAGGVAVSALEMAQNSSRRAWTFEEVDQELQRIMKTIFVNASEAADEFGDSG 425
             F PAKAANAGGV+VSALEMAQNS+R AWTFEEVD +L  IMK I+  + +AA+ +  SG
Sbjct: 375 LFAPAKAANAGGVSVSALEMAQNSTRLAWTFEEVDAKLYEIMKNIYRESIKAAELYEASG 434

Query: 426 NLVLGANIAGFLKVAQAMSAQGIV 449
            NLV+GANIAGF+KVA AM + G+V
Sbjct: 435 NLVVGANIAGFVKVADAMISHGVV 458
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1943

A DNA sequence (GBSx2052) was identified in *S.agalactiae* <SEQ ID 6023> which encodes the amino acid sequence <SEQ ID 6024>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.55   Transmembrane   61 -  77 (  55 -  87)
   INTEGRAL    Likelihood = -7.70   Transmembrane  177 - 193 ( 175 - 202)
   INTEGRAL    Likelihood = -7.06   Transmembrane   99 - 115 (  95 - 122)
   INTEGRAL    Likelihood = -5.89   Transmembrane   42 -  58 (  40 -  60)
   INTEGRAL    Likelihood = -3.08   Transmembrane  160 - 176 ( 159 - 176)
   INTEGRAL    Likelihood = -2.44   Transmembrane  124 - 140 ( 122 - 144)

----- Final Results -----
               bacterial membrane --- Certainty=0.4418(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9955> which encodes amino acid sequence <SEQ ID 9956> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1944

A DNA sequence (GBSx2053) was identified in *S.agalactiae* <SEQ ID 6025> which encodes the amino acid sequence <SEQ ID 6026>. This protein is predicted to be ABC transporter, ATP-binding protein (msbA). Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-10.72   Transmembrane  152 - 168 ( 147 - 192)
   INTEGRAL    Likelihood = -5.47   Transmembrane  267 - 283 ( 264 - 288)
   INTEGRAL    Likelihood = -4.30   Transmembrane  171 - 187 ( 169 - 192)
   INTEGRAL    Likelihood = -2.13   Transmembrane   67 -  83 (  67 -  83)
   INTEGRAL    Likelihood = -0.32   Transmembrane  493 - 509 ( 493 - 509)

----- Final Results -----
               bacterial membrane --- Certainty=0.5288(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

```
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB69752 GB:AL137187 putative ABC transporter [Streptomyces coelicolor A3(2)]
 Identities = 269/611 (44%), Positives = 392/611 (64%), Gaps = 31/611 (5%)

Query: 9   RLWSYLTRYKATLFLAIFLKVLSSFMSILEPFILGLAITELTANLV--DMAKG------- 59
           RL S     +ATLF  +   V+S  ++++ P ILG A    + A +V  DM  G
Sbjct: 27  RLVSQFRPERATLFTLLACVVVSVGLNVVGPKILGRATDLVFAGIVGRDMPSGATKEQVL 86

Query: 60  --------------------VSGAELNVPYIAGILIIYFFRGVFYELGSYGSNYFMTTVV 99
                               V G  ++   +  +L++         L    +   +   V
Sbjct: 87  ATMREHGDGNVADMLRSTDFVPGQGIDFGAVGEVLLLALATFAVAGLLMAVATRLVNRAV 146

Query: 100 QKSIRDIRHDLNRKINKVPVSYFDKHQFGDMLGRFTSDVETVSNALQQSFLQIINAFLSI 159
            +++  +R D+  K++++P+SYFDK Q G++L R T+D++ +   LQQS  Q+IN+ L+I
Sbjct: 147 NRTMFRLREDVQTKLSRLPLSYFDKRQRGEVLSRATNDIDNIGQTLQQSMGQLINSLLTI 206

Query: 160 ILVVVMVLYLNVPLAMIIIACIPVTYFSAQAILKRSQPYFKEQAKILGELNGFVQEKLTG 219
           I V+ M+ Y++  LA++ +  +P+++  A  + KRSQP F +Q +  G+LN  ++E  TG
Sbjct: 207 IGVLAMMFYVSWILALVALVTVPLSFVVATRVGKRSQPQFVQQWRSTGQLNAHIEEMYTG 266

Query: 220 FNIIKLYGREEASSQEFRDITDNLRHVGFKASFISGIMMPVLNSISDFIYLIIAFVGGLQ 279
             ++K++GR+E S+++F +  D L   GFKA F SGIM P++  +S+  Y+++A VGGL+
Sbjct: 267 HALVKVFGRQEESAKQFAEQNDALYEAGFKAQFNSGIMQPLMMCVSNLNYVLVAVVGGLR 326

Query: 280 VIAGTLTIGNMQAFVQYVWQISQPVQTITQLAGVLQSAKSSLERIFEVLD-EEEEANQVT 338
           V +G L+IG++QAF+QY  Q S P+  +  +A ++QS  +S ER+FE+LD EE+ A+ +
Sbjct: 327 VASGQLSIGDVQAFIQYSRQFSMPLTQVASMANLVQSGVASAERVFELLDAEEQSADPIP 386

Query: 339 EKLSHDLTGQVSFHGVDFHYSPDKPLIRDFNLDVEPGQMIAIVGPTGAGKTTLINLLMRF 398
                DL G+V    V F Y P+KPLI D +L VEPG  +AIVGPTGAGKTTL+NLLMRF
Sbjct: 387 GARPEDLRGRVELEHVSFRYDPEKPLIEDLSLKVEPGHTVAIVGPTGAGKTTLVNLLMRF 446

Query: 399 YDVSEGAITVDGHDIRHLSRQDFRQQFGMVLQDAWLYEGTIKENLRFG-NLEASDEDIVA 457
           Y+VS G IT+DG DI  +SR + R   GMVLQD WL+ GTI EN+ +G + E +  +I
Sbjct: 447 YEVSGGRITLDGVDIAKMSRDELRAGIGMVLQDTWLFGGTIAENIAYGASREVTRGEIEE 506

Query: 458 AAKAANVDHFIRTLPGGYNMVMNQESSNISLGQKQLLTIARALLADPKILILDEATSSVD 517
           AA+AA+ D F+RTLP GY+ V++ E + +S G+KQL+TIARA L+DP IL+LDEATSSVD
Sbjct: 507 AARAAHADRFVRTLPDGYDTVIDDEGTGVSAGEKQLITIARAFLSDPVILVLDEATSSVD 566

Query: 518 TRLELLIQKAMKKLMEGRTSFVIAHRLSTIQEADNILVLKDGQIIEQGNHQKLLADKGFY 577
           TR E+LIQKAM KL  GRTSFVIAHRLSTI++AD ILV++DG I+EQG H +LL   G Y
Sbjct: 567 TRTEVLIQKAMAKLAHGRTSFVIAHRLSTIRDADTILVMEDGAIVEQGAHTELLTADGAY 626

Query: 578 YELYNSQFSNS 588
             LY +QF+ +
Sbjct: 627 ARLYKAQFAEA 637
```

There is also homology to SEQ IDs 160 and 6546.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1945

A DNA sequence (GBSx2054) was identified in *S.agalactiae* <SEQ ID 6027> which encodes the amino acid sequence <SEQ ID 6028>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-10.88    Transmembrane  242 - 258 ( 235 - 263)
   INTEGRAL    Likelihood = -9.82    Transmembrane  159 - 175 ( 129 - 177)
   INTEGRAL    Likelihood = -9.71    Transmembrane   52 -  68 (  49 -  77)
```

```
     INTEGRAL    Likelihood = -8.49   Transmembrane  134 - 150 ( 129 - 158)
     INTEGRAL    Likelihood = -1.17   Transmembrane  272 - 288 ( 272 - 289)

----- Final Results -----
               bacterial membrane --- Certainty=0.5352(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB69751 GB:AL137187 putative ABC transporter [Streptomyces
           coelicolor A3(2)]
 Identities = 226/565 (40%), Positives = 342/565 (60%), Gaps = 1/565 (0%)

Query: 6   SYLKRYPNWLWLDLLGAMLFVTVILGMPTALAGMIDNGVTKGDRTGVYLWTFIMFIFVVL 65
           +YL+ Y   + L +    L     L +PT  A +ID GV KGD   +  +  +M    +
Sbjct: 8   TYLRPYKKPIALLVALQFLQTCASLYLPTLNAHIIDEGVVKGDSGYILSYGALMIGISLA 67

Query: 66  GIIGRITMAYASSRLTTTMIRDMRNDMYAKLQEYSHHEYEQIGVSSLVTRMTSDTFVLMQ 125
            ++  I   +  +R    + RD+R  ++ ++Q +S  E    G  SL+TR T+D   +
Sbjct: 68  QVVCNIGAVFYGARTAAALGRDVRGAVFDRVQSFSAREVGHFGAPSLITRTTNDVQQVQM 127

Query: 126 FAEMSLRLGLVTPMVMIFSVVMILITSPSLAWLVAVAMPLLVGVILYVAIKTKPLSERQQ 185
            A M+  L +  P++ +  +VM L     L+ ++   +P+L   +  +  K +PL  + Q
Sbjct: 128 LALMTFTLMVSAPIMCVGGIVMALGLDVPLSGVLLGVVPVLAICVTLIVRKLRPLFRKMQ 187

Query: 186 TMLDKINQYVRENLTGLRVVRAFARENFQSQKFQVANQRYTDTSTGLFKLTGLTEPLFVQ 245
             LD +N+ +RE +TG RV+RAF R+ ++ Q+F+ AN   T+ + G   L  L  P+ +
Sbjct: 188 VRLDTVNRVLREQITGNRVIRAFVRDEYEQQRFRKANTELTEVALGTGNLLALMFPVVMT 247

Query: 246 IIIAMIVAIVWFALDPLQRGAIKIGDLVAFIEYSFHALFSFLLFANLFTMYPRMVVSSHR 305
           ++    +A+VWF    +  G ++IGDL AF+ Y    + S ++    +F M PR  V + R
Sbjct: 248 VVNLSSIAVVWFGAHRIDSGGMQIGDLTAFLAYLMQIVMSVMMATFMFMMVPRAEVCAER 307

Query: 306 IREVMDMPISINPNTEGVTDTKLKGHLEFDNVTFAYPGETESPVLHDISFKAKPGETIAF 365
           I+EV++   S+ P    VT+ +  GHLE     F YPG  E PVL  I   A+PGET A
Sbjct: 308 IQEVLETESSVVPPVAPVTELRRHGHLEIREAGFRYPG-AEEPVLRHIDLVARPGETTAV 366

Query: 366 IGSTGSGKSSLVNLIPRFYDVTLGKILVDGVDVRDYNLKSLRQKIGFIPQKALLFTGTIG 425
           IGSTGSGKS+L+ L+PR +D T G++LV+GVDVR  + K+L + +  +PQK  LF GT+
Sbjct: 367 IGSTGSGKSTLLGLVPRLFDATDGEVLVNGVDVRTVDPKTLAKVVSLVPQKPYLFAGTVA 426

Query: 426 ENLKYGKADATIDDLRQAVDISQAKEFIESHQEAFETHLAEGGSNLSGGQKQRLSIARAV 485
            NL+YG  DAT ++L  A+ ++QAKEF+   +   +  +A+GG+N+SGGQ+QRL+IAR +
Sbjct: 427 TNLRYGNPDATDEELWHALAVAQAKEFVSELEGGLDAPIAQGGTNVSGGQRQRLAIARTL 486

Query: 486 VKDPDLYIFDDSFSALDYKTDATLRARLKEVTGDSTVLIVAQRVGTIMDADQIIVLDEGE 545
           V+ P++Y+FDDSFSALDY TDA LRA L + T ++TV+IVAQRV TI DAD+I+VLDEG
Sbjct: 487 VQRPEIYLFDDSFSALDYATDAALRAELAQETAEATVVIVAQRVATIRDADRIVVLDEGR 546

Query: 546 IVGRGTHAQLIENNAIYREIAESQL 570
           +VG G H +L+ +N  YREI  SQL
Sbjct: 547 VVGVGRHHELMADNETYREIVLSQL 571
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4985> which encodes the amino acid sequence <SEQ ID 4986>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-16.24   Transmembrane  155 - 171 ( 145 - 176)
   INTEGRAL    Likelihood = -7.48   Transmembrane  130 - 146 ( 122 - 150)
   INTEGRAL    Likelihood = -5.04   Transmembrane   13 -  29 (  12 -  30)
   INTEGRAL    Likelihood = -5.04   Transmembrane   56 -  72 (  52 -  75)
   INTEGRAL    Likelihood = -4.14   Transmembrane  239 - 255 ( 238 - 259)
   INTEGRAL    Likelihood = -1.70   Transmembrane  269 - 285 ( 269 - 288)

----- Final Results -----
```

```
                bacterial membrane --- Certainty=0.7496(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 175/511 (34%), Positives = 296/511 (57%), Gaps = 3/511 (0%)

Query: 59  MFIFVVLGIIGRITMAYASSRLTTTMIRDMRNDMYAKLQEYSHHEYEQIGVSSLVTRMTS 118
           + I  +LG++        ++++   +  DMR   + K+Q++S+   E     +LV R+T+
Sbjct: 56  LLIIALLGLMSGAINTVLAAKIAQGVSADMREKTFRKIQDFSYANIEAFNAGNLVVRLTN 115

Query: 119 DTFVLMQFAEMSLRLGLVTPMVMIFSVVMILITSPSLAWLVAVAMPLLVGVILYVAIKTK 178
           D   +     M ++     P++ I + +M + T P L W++ V + L+  ++  V  +
Sbjct: 116 DINQIQSLVMMMFQILFRLPILFIGAFIMAVQTFPQLWWVIVVMVILIALIMGLVMRQMG 175

Query: 179 PLSERQQTMLDKINQYVRENLTGLRVVRAFARENFQSQKFQVANQRYTDTSTGLFKLTGL 238
           P   + Q ++DKIN+  +ENL G+RVV++F +E  Q  KF+  +      +  +     L
Sbjct: 176 PRFGKFQRLMDKINRIAKENLRGVRVVKSFVQEQQQYTKFKETSNDLLALNLSIGYGFSL 235

Query: 239 TEPLFVQIIIAMIVAIVWFALDPLQRGAIKIGDLVAFIEYSFHALFSFLLFANLFTMYPR 298
            +P  + +    +   +      ++     IG++ +F+ Y    +FS ++  ++     R
Sbjct: 236 MQPALMLVSYLAVYVSINVVSTMVETDPTVIGNIASFMTYMMQIMFSIIVVGSMGMQVSR 295

Query: 299 MVVSSHRIREVMDMPISINPNTEGVTDTKLKGHLEFDNVTFAYPGETESPVLHDISFKAK 358
             VS  RIR+++    ++    E   +  + G + FD+V+F YP + E P L  ISF  +
Sbjct: 296 AFVSMARIRQILSTEPAMTFENE--KEETISGSIVFDDVSFTYPNDDE-PTLKHISFAIE 352

Query: 359 PGETIAFIGSTGSGKSSLVNLIPRFYDVTLGKILVDGVDVRDYNLKSLRQKIGFIPQKAL 418
           PG+ +  +G+TGSGKS+L  LIPR +D   G+IL+ G  ++  +  +LRQ +  + QKA+
Sbjct: 353 PGQMVGIVGATGSGKSTLAQLIPRLFDPQDGQILLGGKPIKTLSQTTLRQSVSIVLQKAI 412

Query: 419 LFTGTIGENLKYGKADATIDDLRQAVDISQAKEFIESHQEAFETHLAEGGSNLSGGQKQR 478
           LF+GTI +NL+ G A A ID +++A  I+QAKEFI+     +E+ + E GSNLSGGQKQR
Sbjct: 413 LFSGTIADNLRQGSAKADIDAMQKAAQIAQAKEFIDRMDSRYESQVEERGSNLSGGQKQR 472

Query: 479 LSIARAVVKDPDLYIFDDSFSALDYKTDATLRARLKEVTGDSTVLIVAQRVGTIMDADQI 538
           LSIAR V+  P + I DDS SALD K++  ++  L      +T +IVAQ++ +++ AD+I
Sbjct: 473 LSIARGVINHPKILILDDSTSALDAKSEKRVQEALSHKLEGTTTVIVAQKISSVVKADKI 532

Query: 539 IVLDEGEIVGRGTHAQLIENNAIYREIAESQ 569
           +VLD+G+++G GTHA+L+ NNAIYREI E+Q
Sbjct: 533 LVLDQGQLIGEGTHAELVANNAIYREIYETQ 563
```

There is also homology to SEQ IDs 72 and 6552.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1946

A DNA sequence (GBSx2055) was identified in *S.agalactiae* <SEQ ID 6029> which encodes the amino acid sequence <SEQ ID 6030>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2391(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA51784 GB:X73368 ORF 18.3 [Salmonella typhimurium]
```

```
Identities = 58/162 (35%), Positives = 92/162 (55%), Gaps = 8/162 (4%)

Query: 1   MIIRPIIKNDDQAVAQLIRQSLRAYDL--DKPDTAYSDPHLDHLTSYYEKIEKSGFFVIE 58
           + +R I   D+ A+A++IRQ    Y L  DK  T  +DP+LD L   Y +   + ++V+E
Sbjct: 9   LTVRRITTADNAAIARVIRQVSAEYGLTADKGYTV-ADPNLDELYQVYSQ-PGAAYWVVE 66

Query: 59  ERDEIIGCGGFGPLKNL---IAEMQKVYIAERFRGKGLATDLVKMIEVEARKIGYRQLYL 115
           +   ++G GG  PL      I E+QK+Y     RG+GLA  L  M    AR+ G+++ YL
Sbjct: 67  QNGCVVGGGGVAPLSCSEPDICELQKMYFLPVIRGQGLAKKLALMALDHAREQGFKRCYL 126

Query: 116 ETASTLSRATAVYKHMGYCALSQPIANDQGHTAMDIWMIKDL 157
           ET + L  A A+Y+ +G+  +S+P+    GH   ++ M+KDL
Sbjct: 127 ETTAFLREAIALYERLGFEHISEPL-GCTGHVDCEVRMLKDL 167
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1947

A DNA sequence (GBSx2056) was identified in *S.agalactiae* <SEQ ID 6031> which encodes the amino acid sequence <SEQ ID 6032>. This protein is predicted to be ABC transporter. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1738(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12566 GB:Z99108 similar to ABC transporter (ATP-binding
           protein) [Bacillus subtilis]
 Identities = 269/625 (43%), Positives = 397/625 (63%), Gaps = 11/625 (1%)

Query: 1   MSDFLVDGLTKSVGDKTVFSNVSFIIHSLDRIGIIGVNGTGKTTLLDVISGELGFDGDRS 60
           MS    + L K+ GDKT+F ++SF I    +RIG+IG NGTGK+TLL VI+G    +
Sbjct: 1   MSILKAENLYKTYGDKTLFDHISFHIEENERIGLIGPNGTGKSTLLKVIAGLESIE--EG 58

Query: 61  PFSSANDYKIAYLKQEPDFDDSQTILDTVLSSDLREMALIKEYELLLNHY-----EESKQ 115
             + +   ++ +L Q+P+    QT+L+ + S +   M  ++EYE  L        E  +Q
Sbjct: 59  EITKSGSVQVEFLHQDPELPAGQTVLEHIYSGESAVMKTLREYEKALYELGKDPENEQRQ 118

Query: 116 SRLEKVMAEMDSLDAWSIESEVKTVLSKLGITDLQLSVGELSGGLRRRVQLAQVLLNDAD 175
             L    A+MD+ +AW   +  KTVLSKLG+ D+   V ELSGG ++RV +A+ L+  AD
Sbjct: 119 KHLLAAQAKMDANNAWDANTLAKTVLSKLGVNDVTKPVNELSGGQKKRVAIAKNLIQPAD 178

Query: 176 LLLLDEPTNHLDIDTIAWLTNFLKNSKKTVLFITHDRYFLDNVATRIFELDKAQITEYQG 235
           LL+LDEPTNHLD +TI WL  +L      V+ +THDRYFL+ V  RI+EL++  +  Y+G
Sbjct: 179 LLILDEPTNHLDNETIEWLEGYLSQYPGAVMLVTHDRYFLNRVTNRIYELERGSLYTYKG 238

Query: 236 NYQDYVRLRAEQDERDAASLHKKKQLYKQELAWMRTQPQARATKQQARINRFQNLKNDLH 295
           NY+ ++  RAE++ +      K++ L ++ELAW+R   +AR+TKQ+ARI+R + LK
Sbjct: 239 NYEVFLEKRAEREAQAEQKETKRQNLLRRELAWLRRGAKARSTKQKARIDRVETLKEQTG 298

Query: 296 QTSDTSDLEMTFETSRIGKKVINFENVSFSYPDKSILKDFNLLIQNKDRIGIVGDNGVGK 355
             S  S L+    + R+GK+VI  ENV  +Y  + ++  FN L+    +RIGI+G NG+GK
Sbjct: 299 PQSSGS-LDFAIGSHRLGKQVIEAENVMIAYDGRMLVDRFNELVIPGERIGIIGPNGIGK 357

Query: 356 STLLNLIVQDLQPDSGNVSIGETIRVGYFSQQLHNMDGSKRVINYLQEVADEVKTSVGTT 415
           +TLLN +     PD G+++IG+T+R+GY++Q    M+G  +VI+Y++E A+ VKT+ G
```

```
Sbjct: 358 TTLLNALAGRHTPDGGDITIGQTVRIGYYTQDHSEMNGELKVIDYIKETAEVVKTADGDM 417

Query: 416 SVTE-LLEQFLFPRSTHGTQIAKLSGGEKKRLYLLKILIEKPNVLLLDEPTNDLDIATLT 474
               E +LE+FLFPRS   T I KLSGGEK+RLYLL++L+++PNVL LDEPTNDLD  TL+
Sbjct: 418 ITAEQMLERFLFPRSMQQTYIRKLSGGEKRRLYLLQVLMQEPNVLFLDEPTNDLDTETLS 477

Query: 475 VLENFLQGFGGPVITVSHDRYFLDKVANKIIAFEDND-IREFFGNYTDYLDEKAFNEQNN 533
            VLE+++  F G VITVSHDRYFLD+V +++I FE N  I  F G+Y+DY++E    +
Sbjct: 478 VLEDYIDQFPGVVITVSHDRYFLDRVVDRLIVFEGNGVISRFQGSYSDYMEESKAKKAAP 537

Query: 534 EVISKKESTKTSREKQSRKRMSYFEKQEWATIEDDIMILENTITRIENDMQTCGSDFTRL 593
            +  + +E T  +  K+ RK++SY ++ EW  IED I  LE    ++E D+   GSDF ++
Sbjct: 538 KP-AAEEKTAEAEPKKKRKKLSYKDQLEWDGIEDKIAQLEEKHEQLEADIAAAGSDFGKI 596

Query: 594 SDLQKELDAKNEALLEKYDRYEYLS 618
             +L  E     E L    DR+  LS
Sbjct: 597 QELMAEQAKTAEELEAAMDRWTELS 621
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6033> which encodes the amino acid sequence <SEQ ID 6034>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2591(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 467/624 (74%), Positives = 535/624 (84%), Gaps = 3/624 (0%)

Query: 1   MSDFLVDGLTKSVGDKTVFSNVSFIIHSLDRIGIIGVNGTGKTTLLDVISGELGFDGDRS 60
            MS FLV+ LTK+VGDKTVF ++SFIIH  DRIGIIGVNGTGKTTLLDV+SG LGFDGD S
Sbjct: 1   MSHFLVEKLTKTVGDKTVFQDISFIIHDFDRIGIIGVNGTGKTTLLDVLSGRLGFDGDHS 60

Query: 61  PFSSANDYKIAYLKQEPDFDDSQTILDTVLSSDLREMALIKEYELLLNHYEESKQSRLEK 120
            PFS ANDYKIAYL Q+P+F+D+ ++LDTVLS+D++ + LI++YELL+ +Y E KQ  LE
Sbjct: 61  PFSKANDYKIAYLTQDPEFNDAASVLDTVLSADVKAIQLIRQYELLMANYTEDKQESLES 120

Query: 121 VMAEMDSLDAWSIESEVKTVLSKLGITDLQLSVGELSGGLRRRVQLAQVLLNDADLLLLD 180
            +M+EMD LDAWSIES+VKTVLSKLGITDL+  VG+LSGG+RRRVQLAQVLL  ADLLLLD
Sbjct: 121 LMSEMDRLDAWSIESDVKTVLSKLGITDLEQKVGDLSGGMRRRVQLAQVLLGAADLLLLD 180

Query: 181 EPTNHLDIDTIAWLTNFLKNSKKTVLFITHDRYFLDNVATRIFELDKAQITEYQGNYQDY 240
            EPTNHLDIDTIAWLT +LK +KKTVLFITHDRYFLD+VATRIFELDKA +TEYQGNYQDY
Sbjct: 181 EPTNHLDIDTIAWLTTYLKTAKKTVLFITHDRYFLDHVATRIFELDKAGLTEYQGNYQDY 240

Query: 241 VRLRAEQDERDAASLHKKKQLYKQELAWMRTQPQARATKQQARINRFQNLKNDLHQTSDT 300
            VRL+AEQDERDAA+LHKKKQLYKQELAWMRTQPQARATKQQARINRF +LK ++HQ S
Sbjct: 241 VRLKAEQDERDAANLHKKKQLYKQELAWMRTQPQARATKQQARINRFSDLKKEVHQDSSA 300

Query: 301 SDLEMTFETSRIGKKVINFENVSFSYPDKSILKDFNLLIQNKDRIGIVGDNGVGKSTLLN 360
              LEMTFETSRIGKKVI+FE++SF+Y D+ ++KDFNL+IQNKDRIGIVGDNGVGKSTLLN
Sbjct: 301 DKLEMTFETSRIGKKVIHFEDLSFAYGDRQLIKDFNLIIQNKDRIGIVGDNGVGKSTLLN 360

Query: 361 LIVQDLQPDSGNVSIGETIRVGYFSQQLHNMDGSKRVINYLQEVADEVKTSVGTTSVTEL 420
            +I  DL+P SG + IG+TIRVGYFSQQL +MD +KRVINYLQEVADEVKTSVGTTS++EL
Sbjct: 361 IINGDLKPTSGKLDIGDTIRVGYFSQQLKDMDETKRVINYLQEVADEVKTSVGTTSISEL 420

Query: 421 LEQFLFPRSTHGTQIAKLSGGEKKRLYLLKILIEKPNVLLLDEPTNDLDIATLTVLENFL 480
            LEQFLFPRS+HGT IAKLSGGEKKRLYLLK+LIEKPNVLLLDEPTNDLDIATL VLENFL
Sbjct: 421 LEQFLFPRSSHGTLIAKLSGGEKKRLYLLKLLIEKPNVLLLDEPTNDLDIATLKVLENFL 480

Query: 481 QGFGGPVITVSHDRYFLDKVANKIIAFEDNDIREFFGNYTDYLDEKAFNEQNNEVISKKE 540
              F GPVITVSHDRYFLDKVA KI+AFE+ DIR F+GNY+DYLDEK F ++  E    K
```

```
Sbjct: 481 ANFAGPVITVSHDRYFLDKVATKILAFEEGDIRVFYGNYSDYLDEKVFEKETVEADLAKT 540

Query: 541 STKTS---REKQSRKRMSYFEKQEWATIEDDIMILENTITRIENDMQTCGSDFTRLSDLQ 597
           +        +K+ RKRMSY EKQEWA IED I  +E  I  IEN M T  SD+ +L+ LQ
Sbjct: 541 TVTEEVPLPQKEERKRMSYLEKQEWAQIEDKIATIEANIEEIENQMLTVVSDYGQLAQLQ 600

Query: 598 KELDAKNEALLEKYDRYEYLSELD 621
           KELD +N  LL  Y+R+EYLS LD
Sbjct: 601 KELDQRNNDLLLAYERFEYLSGLD 624
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1948

A DNA sequence (GBSx2057) was identified in *S.agalactiae* <SEQ ID 6035> which encodes the amino acid sequence <SEQ ID 6036>. This protein is predicted to be poly(a) polymerase (papS). Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2658(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9957> which encodes amino acid sequence <SEQ ID 9958> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB38446 GB:L47709 poly(A) polymerase [Bacillus subtilis]
 Identities = 157/395 (39%), Positives = 235/395 (58%), Gaps = 14/395 (3%)

Query: 11  FQKALPILKKIKKAGYEAYFVGGSVRDVLLDRPIHDVDIATSSYPEETKQIFKRTVDVGI 70
            F KALP+L+ + +AG++AYFVGG+VRD  + R I DVDIAT + P++ +++F+RTVDVG
Sbjct: 5   FIKALPVLRILIEAGHQAYFVGGAVRDSYMKRTIGDVDIATDAAPDQVERLFQRTVDVGK 64

Query: 71  EHGTVLVLEKGGEYEITTFRTEEVYVDYRRPSQVNFVRSLEEDLKRRDFTVNAFALNEDG 130
            EHGT++VL +   YE+TTFRTE  YVD+RRPS+V F+ SLEEDLKRRD T+NA A+  DG
Sbjct: 65  EHGTIIVLWEDETYEVTTFRTESDYVDFRRPSEVQFISSLEEDLKRRDLTINAMAMTADG 124

Query: 131 EVIDLFHGLDDLDNHLLRAVGLASERFNEDALRIMRGLRFSASLNFDIETTTFEAMKKHA 190
           +V+D F G  D+D  ++R VG   +RF EDALR++R +RF + L F +   T EA+ K
Sbjct: 125 KVLDYFGGKKDIDQKVIRTVGKPEDRFQEDALRMLRAVRFMSQLGFTLSPETEEAIAKEK 184

Query: 191 SLLEKISVERSFIEFDKLLLAPYWRKGMLALIDSHAFNYLPCLKNRELQLSAFLSQLDKD 250
           SLL  +SVER  IEF+KLL     R+ +  LI +  + LP   ++   L   +S  +
Sbjct: 185 SLLSHVSVERKTIEFEKLLQGRASRQALQTLIQTRLYEELPGFYHKRENL---ISTSEFP 241

Query: 251 FLFETS-EQAWASLILSMEV--EHTKTFLKKWKTSTHFQKDVEHIVDVYRIREQMGLTKE 307
           F   TS E+ WA+L++++ +  +     FLK WK      K+  HI D +       L
Sbjct: 242 FFSLTSREELWAALLINLGIVLKDAPLFLKAWKLPGKVIKEAIHIADTF----GQSLDAM 297

Query: 308 HLYRYGKTIIKQAEGIRKAR-GLMVDFEKIEQLD---SELAIHDRHEIVVNGGTLIKKLG 363
            +YR GK  +  A  I + R    +D +K++ +      L I    ++ + G  L+
Sbjct: 298 TMYRAGKKALLSAAKISQLRQNEKLDEKKLKDIQYAYQNLPIKSLKDLDITGKDLLALRN 357

Query: 364 IKPGPQMGDIISQIELAIVLGQLINEEEAILHFVK 398
              G  + + +  IE A+V G+L N+++ I  ++K
Sbjct: 358 RPAGKWVSEELQWIEQAVVTGKLSNQKKHIEEWLK 392
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6037> which encodes the amino acid sequence <SEQ ID 6038>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2023(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 256/400 (64%), Positives = 312/400 (78%)

Query: 2   MRLNYLPSEFQKALPILKKIKKAGYEAYFVGGSVRDVLLDRPIHDVDIATSSYPEETKQI 61
           M+L  +PSEFQKALPIL KIK+AGYEAYFVGGSVRDVLL+RPIHDVDIATSSYPEETK I
Sbjct: 1   MKLMTMPSEFQKALPILTKIKEAGYEAYFVGGSVRDVLLERPIHDVDIATSSYPEETKAI 60

Query: 62  FKRTVDVGIEHGTVLVLEKGGEYEITTFRTEEVYVDYRRPSQVNFVRSLEEDLKRRDFTV 121
           F RTVDVGIEHGTVLVLE GGEYEITTFRTE++YVDYRRPSQV+FVRSLEEDLKRRDFTV
Sbjct: 61  FNRTVDVGIEHGTVLVLENGGEYEITTFRTEDIYVDYRRPSQVSFVRSLEEDLKRRDFTV 120

Query: 122 NAFALNEDGEVIDLFHGLDDLDNHLLRAVGLASERFNEDALRIMRGLRFSASLNFDIETT 181
           NA AL+E+G+VID F GL DL    LRAVG A ERF EDALRIMRG RF+ASL+FDIE
Sbjct: 121 NALALDENGQVIDKFRGLIDLKQKRLRAVGKAEERFEEDALRIMRGFRFAASLDFDIEAI 180

Query: 182 TFEAMKKHASLLEKISVERSFIEFDKLLLAPYWRKGMLALIDSHAFNYLPCLKNRELQLS 241
           TFEAM+ H+ LLEKISVERSF EFDKLL+AP+WRKG+ A+I   A++YLP LK +E  L+
Sbjct: 181 TFEAMRSHSPLLEKISVERSFTEFDKLLMAPHWRKGISAMIACQAYDYLPGLKQQEAGLN 240

Query: 242 AFLSQLDKDFLFETSEQAWASLILSMEVEHTKTFLKKWKTSTHFQKDVEHIVDVYRIREQ 301
            +   L  +F F    QAWA +++S+ +E  K+FLK WKTS  FQ+ V  ++ +YRIR++
Sbjct: 241 HLIVSLKDNFTFSDYHQAWAYVMISLAIEDPKSFLKAWKTSNDFQRYVTKLIALYRIRQE 300

Query: 302 MGLTKEHLYRYGKTIIKQAEGIRKARGLMVDFEKIEQLDSELAIHDRHEIVVNGGTLIKK 361
               K  +Y+YGK +    E +RKA+ L VD ++I  LD  L IHD+H+IV+NG  LIK
Sbjct: 301 RSFEKLDIYQYGKKMASLVEDLRKAQSLSVDMDRINTLDQALVIHDKHDIVLNGSHLIKD 360

Query: 362 LGIKPGPQMGDIISQIELAIVLGQLINEEEAILHFVKQYL 401
            G+K GPQ+G ++ ++ELAIV G+L N+   I  FV++ L
Sbjct: 361 FGMKSGPQLGLMLEKVELAIVEGRLDNDFTTIEAFVREEL 400
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1949

A DNA sequence (GBSx2058) was identified in *S.agalactiae* <SEQ ID 6039> which encodes the amino acid sequence <SEQ ID 6040>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2939(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07346 GB:AP001519 unknown conserved protein [Bacillus halodurans]
 Identities = 94/274 (34%), Positives = 153/274 (55%), Gaps = 2/274 (0%)

Query: 2   KLALITDTSAYLPEAIENHEDVYVLDIPIIIDGKTYIEGQNLTLDQYYDKLAASKELPKT 61
```

```
           K+A++TD++AYL         V V+ + ++   + Y E   L+   +Y+KL   ++LP T
Sbjct: 3   KIAIVTDSTAYLGPKRAKELGVIVVPLSVVFGEEAYQEEVELSSADFYEKLKHEEKLPTT 62

Query: 62  SQPSLAELDDLLCQLEKEGYTHVLGLFIAAGISGFWQNIQFLIEEHPNLTIAFPDTKITS 121
           SQP++    +   +L KEG+  V+ + +++ ISG +Q+          + +   D+ I+
Sbjct: 63  SQPAVGLFVETFERLAKEGFEVVISIHLSSKISGTYQSALTAGSMVEGIEVIGYDSGISC 122

Query: 122 APQGNLVRNALMCSREGMDFDVIVNKIQSQIEKIEGFIVVNDLNHLVKGGRLSNGSAIIG 181
            PQ N V  A    +EG D   I++ +    ++     VV+DL+HL +GGRL+    ++G
Sbjct: 123 EPQANFVAEAAKLVKEGADPQTIIDHLDEVKKRTNALFVVHDLSHLHRGGRLNAAQLVVG 182

Query: 182 NLLSIKPVLHFNEEGKIVVYEKVRTEKKALKRLAEI-VKEMTADGEYDIAIIHSRAQDKA 240
           +LL IKP+LHF E+G IV  EKVRTEKKA  R+ E+  +E ++       +IH+   D A
Sbjct: 183 SLLKIKPILHF-EDGSIVPLEKVRTEKKAWARVKELFAEEASSASSVKATVIHANRLDGA 241

Query: 241 EQLYNLLAKAGLKDDLEIVSFGGVIATHLGEGAV 274
           E+L + +       D+ I  FG VI THLGEG++
Sbjct: 242 EKLADEIRSQFSHVDVSISHFGPVIGTHLGEGSI 275
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6041> which encodes the amino acid sequence <SEQ ID 6042>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3379(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 181/281 (64%), Positives = 233/281 (82%)

Query: 1   MKLALITDTSAYLPEAIENHEDVYVLDIPIIIDGKTYIEGQNLTLDQYYDKLAASKELPK 60
           MKLA+ITD++A LP  ++  + ++ LDIP+IID +TY EG+NL++D +Y K+A S+ LPK
Sbjct: 1   MKLAVITDSTATLPTDLKQDKAIFSLDIPVIIDDETYFEGRNLSIDDFYQKMADSQNLPK 60

Query: 61  TSQPSLAELDDLLCQLEKEGYTHVLGLFIAAGISGFWQNIQFLIEEHPNLTIAFPDTKIT 120
           TSQPSL+ELD+LL  L  +GYTHV+GLF+A GISGFWQNIQFL EEHP + +AFPD+KIT
Sbjct: 61  TSQPSLSELDNLLGLLSSKGYTHVIGLFLAGGISGFWQNIQFLAEEHPEIEMAFPDSKIT 120

Query: 121 SAPQGNLVRNALMCSREGMDFDVIVNKIQSQIEKIEGFIVVNDLNHLVKGGRLSNGSAII 180
           SAP G++V+N L  SR+GM F  I+NK+Q QI+    FI+V+DLNHLVKGGRLSNGSA++
Sbjct: 121 SAPLGSMVKNVLDWSRQGMTFQAILNKLQEQIDGTTAFIMVDDLNHLVKGGRLSNGSALL 180

Query: 181 GNLLSIKPVLHFNEEGKIVVYEKVRTEKKALKRLAEIVKEMTADGEYDIAIIHSRAQDKA 240
           GNLLSIKP+L F+EEGKIVVYEKVRTEKKA+KRL EI+ ++ ADG+Y++ IIHS+AQDKA
Sbjct: 181 GNLLSIKPILRFDEEGKIVVYEKVRTEKKAMKRLVEILNDLIADGQYNVFIIHSKAQDKA 240

Query: 241 EQLYNLLAKAGLKDDLEIVSFGGVIATHLGEGAVAFGITPK 281
           + L  LL  +G + D+E V FG VIATHLGEGA+AFG+TP+
Sbjct: 241 DYLKRLLQDSGYQYDIEEVHFGAVIATHLGEGAIAFGVTPR 281
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1950

A DNA sequence (GBSx2059) was identified in *S.agalactiae* <SEQ ID 6043> which encodes the amino acid sequence <SEQ ID 6044>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.59   Transmembrane   51 -  67 (  50 -  67)
```

```
----- Final Results -----
               bacterial membrane --- Certainty=0.1638(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6045> which encodes the amino acid sequence <SEQ ID 6046>. Analysis of this protein sequence reveals the following:

```
     Possible site: 61

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.19    Transmembrane   50 -  66 (  49 -  67)

----- Final Results -----
               bacterial membrane --- Certainty=0.2275(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 94/126 (74%), Positives = 115/126 (90%)

Query: 1   MEVIREQEFVNQYHYDARNLEWEEENGTPKTNFEVTFQLANRDEAAKVTSIVAVLQFVIV 60
           M+++RE+EFVNQYHYDARNLEWE+ENGTP+TNFEVTFQL ++DE  K T IV+VLQFVIV
Sbjct: 1   MQLVREKEFVNQYHYDARNLEWEKENGTPETNFEVTFQLIDKDEQQKETVIVSVLQFVIV 60

Query: 61  RDEFVISGVISQMAHIQGRLINEPSEFSQDEVENLAAPLLEIVKRLTYEVTEIALDRPGV 120
           ++EFVISGVISQM  I  RL+++PSEF+Q+EVE+LAAPLL++VKRLTYEVTEIALDRPG+
Sbjct: 61  KEEFVISGVISQMVRILDRLVDKPSEFTQEEVESLAAPLLDMVKRLTYEVTEIALDRPGI 120

Query: 121 TLEFNS 126
            LEF +
Sbjct: 121 HLEFKN 126
```

SEQ ID 6044 (GBS416) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 79 (lane 4; MW 17.5kDa).

GBS416-His was purified as shown in Figure 214, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1951

A DNA sequence (GBSx2060) was identified in *S.agalactiae* <SEQ ID 6047> which encodes the amino acid sequence <SEQ ID 6048>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3875(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1952

A DNA sequence (GBSx2061) was identified in *S.agalactiae* <SEQ ID 6049> which encodes the amino acid sequence <SEQ ID 6050>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1953

A DNA sequence (GBSx2062) was identified in *S.agalactiae* <SEQ ID 6051> which encodes the amino acid sequence <SEQ ID 6052>. This protein is predicted to be PTS system, fructose-specific enzyme II, BC component (fruA-1). Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.56   Transmembrane  630 - 646  ( 618 - 653)
   INTEGRAL    Likelihood = -7.43   Transmembrane  307 - 323  ( 303 - 331)
   INTEGRAL    Likelihood = -7.01   Transmembrane  415 - 431  ( 412 - 435)
   INTEGRAL    Likelihood = -7.01   Transmembrane  448 - 464  ( 444 - 474)
   INTEGRAL    Likelihood = -3.72   Transmembrane  595 - 611  ( 591 - 612)
   INTEGRAL    Likelihood = -3.61   Transmembrane  530 - 546  ( 529 - 553)
   INTEGRAL    Likelihood = -2.39   Transmembrane  350 - 366  ( 350 - 371)
   INTEGRAL    Likelihood = -1.70   Transmembrane  486 - 502  ( 486 - 506)
   INTEGRAL    Likelihood = -1.49   Transmembrane  376 - 392  ( 376 - 392)

----- Final Results -----
               bacterial membrane --- Certainty=0.5225(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9959> which encodes amino acid sequence <SEQ ID 9960> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04547 GB:AP001510 PTS system, fructose-specific enzyme II, BC
          component [Bacillus halodurans]
 Identities = 320/659 (48%), Positives = 438/659 (65%), Gaps = 46/659 (6%)

Query: 1    MKIQDLLKKEVMIMDLKATSKEAAIDEMITKLVDTGVVTNFAIFKDGIMKREAQTSTGLG 60
            +KI +LLKK+ M+++L+A SKEA IDE++  L   G + +   FK  I++RE+Q++TG+G
Sbjct: 2    LKISELLKKDTMVLNLRAASKEAVIDELVRTLDKAGRLNDAQAFKRAILERESQSTTGVG 61

Query: 61   DGIAMPHSKNAAVKEATVLFAKSASGVDYEALDGQPTDLFFMIAAPDGANDTHLAALAEL 120
            +GIA+PH+K AAVK+  + F +S +G+DYE+LDGQP+ LFFMIAA +GAN+ HL  L+ L
Sbjct: 62   EGIAIPHAKTAAVKQPAIAFGRSDAGIDYESLDGQPSHLFFMIAASEGANNEHLETLSRL 121
```

```
Query: 121 SKYLLKEGFADQLRQAKTPDDIIATFDSNSISQETVAPQTVQSTSKGSDYIVAVTACTTG 180
           S +L+ E F   L +A++ D+I+A  D     +E       +   +G + ++AVT C TG
Sbjct: 122 STFLMDETFRSTLMKAQSEDEILAAID----KKEAETAGEAEEKQEGYE-LLAVTGCPTG 176

Query: 181 IAHTYMAEEALKKKAAEMGVGIKVETNGASGVGNKLTSSDIARAKGVIIAADKAVEMDRF 240
           IAHTYMA + LK KA E+GV IKVETNG+ GV N+LT  +I+ AK +I+AAD  VEMDRF
Sbjct: 177 IAHTYMAADNLKSKAQELGVSIKVETNGSGGVKNRLTDEEISAAKAIIVAADTKVEMDRF 236

Query: 241 DGKPLVSRPVADGIKKSEDLINIILDNKAQTYHAKNQNDKQSGESDGKSGLGS---AFYK 297
            GKP++  PV DGI++ ++LI+  L  KA  Y    +   Q+   DG +G G     FYK
Sbjct: 237 HGKPVIQVPVTDGIRRPKELIDQALAGKAPVY----EGGAQASGEDGSAGGGRPKLGFYK 292

Query: 298 HLMGGVSQMLPFVIGGGIMIAIAFLFDNILGVPKDQLSNLGSYHEIAALFKNIGGA-AFA 356
           HLM GVS MLPFV+GGGI+IAI+F+F      P D      SYH  A +   IGG  AF
Sbjct: 293 HLMNGVSNMLPFVVGGGILIAISFMFGIKAFDPSDP-----SYHPFAEMLMTIGGGNAFG 347

Query: 357 FMLPVLAGYIAYSIAEKPGLVAGFVAGSIASSGLAFGKVPFAEGGKATLALAGVPSGFLG 416
            M+PVLA +IA SIA++PG  AG + G IAS+G A                     GFLG
Sbjct: 348 LMIPVLAAFIAMSIADRPGFAAGMIGGLIASTGEA---------------------GFLG 386

Query: 417 ALVGGFLAGGVILLLRKLLSGLPKSLEGIKSILLYPLLGVLITGFLMLLVNIPMAAINTA 476
            L+ GFLAG V L ++K+L+ LP++L+GIK+IL YP+  + ITG +ML++  P+AA NT
Sbjct: 387 GLIAGFLAGYVALGVKKVLANLPQTLDGIKTILFYPVFNIFITGMIMLVIVGPLAAFNTG 446

Query: 477 LNTFLQGLSGSSAVLMGLLVGGMMAVDMGGPVNKAAYVFGTGTLAATVANGGSVVMAAVM 536
           L  +L  +  ++ V++G+++GGMMAVDMGGP+NKAA+ FG   + A    G     AAVM
Sbjct: 447 LQDWLGSMGTANMVILGVILGGMMAVDMGGPINKAAFTFGIAMIDA----GNFGPHAAVM 502

Query: 537 AGGMVPPLAVFVATLLFKDKFNNEERQSGLTNIVMGLSFITEGAIPFGAADPARAIPSFI 596
           AGGMVPPL + +AT LFK KF  +ER++G TN ++G SFITEGAIPF AADP R IPS I
Sbjct: 503 AGGMVPPLGIALATTLFKKKFTKQEREAGKTNYILGASFITEGAIPFAAADPGRVIPSII 562

Query: 597 VGSALTGALVGLAGIKLMAPHGGIFVI---ALTSNPLLYILFILIGAVVSGVLFGLFRK 652
           VGSA  G L  L  + L APHGG FVI    + +NPLLY++ I+ G++V+ +L G ++K
Sbjct: 563 VGSAFAGGLTALFNVTLSAPHGGAFVIFIGNIVNNPLLYLVAIIAGSIVTALLLGFWKK 621
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6053> which encodes the amino acid sequence <SEQ ID 6054>. Analysis of this protein sequence reveals the following:

```
     Possible site: 18
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.77    Transmembrane   624 - 640 ( 612 - 646)
   INTEGRAL    Likelihood = -7.59    Transmembrane   301 - 317 ( 297 - 321)
   INTEGRAL    Likelihood = -6.85    Transmembrane   442 - 458 ( 439 - 468)
   INTEGRAL    Likelihood = -5.95    Transmembrane   409 - 425 ( 406 - 426)
   INTEGRAL    Likelihood = -3.61    Transmembrane   524 - 540 ( 523 - 547)
   INTEGRAL    Likelihood = -2.50    Transmembrane   337 - 353 ( 337 - 353)
   INTEGRAL    Likelihood = -2.44    Transmembrane   589 - 605 ( 589 - 605)
   INTEGRAL    Likelihood = -1.70    Transmembrane   480 - 496 ( 480 - 500)
   INTEGRAL    Likelihood = -1.44    Transmembrane   370 - 386 ( 370 - 386)

----- Final Results -----
               bacterial membrane --- Certainty=0.5310(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB04547 GB:AP001510 PTS system, fructose-specific enzyme II, BC
           component [Bacillus halodurans]
 Identities = 322/659 (48%), Positives = 431/659 (64%), Gaps = 48/659 (7%)

Query: 1   MKIQDLLRKDIMILDLQAISKEVAIDEMITKLVEKDIVHDFDVFKKSIMTREEQTSTGLG 60
           +KI +LL+KD M+L+L+A SKE  IDE++  L +   ++D   FK++I+ RE Q++TG+G
Sbjct: 2   LKISELLKKDTMVLNLRAASKEAVIDELVRTLDKAGRLNDAQAFKRAILERESQSTTGVG 61
```

```
Query: 61   DGIAMPHSKNIVVDKPAVLFAKSNKGVDYKALDGQPTDLFFMIAAPQGANDTHLAALAEL 120
            +GIA+PH+K   V +PA+ F +S+ G+DY++LDGQP+ LFFMIAA +GAN+ HL  L+ L
Sbjct: 62   EGIAIPHAKTAAVKQPAIAFGRSDAGIDYESLDGQPSHLFFMIAASEGANNEHLETLSRL 121

Query: 121  SQYLLKDGFADKLRAAATPEAVIAVFD--EASTAKEEVVAPTSGQDFIVAVTACPTGIAH 178
            S +L+ + F   L  A + + ++A  D  EA TA E        +  ++AVT CPTGIAH
Sbjct: 122  STFLMDETFRSTLMKAQSEDEILAAIDKKEAETAGEAEEKQEGYE--LLAVTGCPTGIAH 179

Query: 179  TYMAEEALKKQAAEMGVAIKVETNGASGVANRLTAEDIQRAKGVIVAADKAVEMDRFDGK 238
            TYMA + LK +A E+GV+IKVETNG+ GV NRLT E+I  AK +IVAAD  VEMDRF GK
Sbjct: 180  TYMAADNLKSKAQELGVSIKVETNGSGGVKNRLTDEEISAAKAIIVAADTKVEMDRFHGK 239

Query: 239  QFIARPVADGIKKSQELISLILNNEGNTYHAKNGKSETAVSTEKTSLGG-----AFYKHL 293
              I  PV DGI++ +ELI   L  +   Y     +     S E  S GG      FYKHL
Sbjct: 240  PVIQVPVTDGIRRPKELIDQALAGKAPVY-----EGGAQASGEDGSAGGGRPKLGFYKHL 294

Query: 294  MGGVSQMLPFVIGGGIMIALAFLLDNMLGVPNDQLGSLGSYHEIAAIFMNIGGA-AFSFM 352
            M GVS MLPFV+GGGI+IA++F+       P+D      SYH  A + M IGG  AF  M
Sbjct: 295  MNGVSNMLPFVVGGGILIAISFMFGIKAFDPSDP-----SYHPFAEMLMTIGGGNAFGLM 349

Query: 353  LPVLAGYIAYSIAEKPGLVAGFVAGAIASNGLAFGKVPFAAGGEVSLGLTGVPSGFLGAL 412
            +PVLA +IA SIA++PG  AG + G IAS G A                     GFLG L
Sbjct: 350  IPVLAAFIAMSIADRPGFAAGMIGGLIASTGEA---------------------GFLGGL 388

Query: 413  VGGFLAGGVILALRKLLAGLPRSLEGVKSILLYPLLGVLVTGFLMLFVNIPMAAINTALN 472
            + GFLAG V L ++K+LA LP++L+G+K+IL YP+  + +TG +ML +  P+AA NT L
Sbjct: 389  IAGFLAGYVALGVKKVLANLPQTLDGIKTILFYPVFNIFITGMIMLVIVGPLAAFNTGLQ 448

Query: 473  DFLQGLSGSSAVLMGLLVGGMMAVDMGGPVNKAAYVFGTGTLAATVANGGSVVMAAVMAG 532
            D+L  +  ++ V++G+++GGMMAVDMGGP+NKAA+ FG   + A    G     AAVMAG
Sbjct: 449  DWLGSMGTANMVILGVILGGMMAVDMGGPINKAAFTFGIAMIDA----GNFGPHAAVMAG 504

Query: 533  GMVPPLAVFVATLLFKDKFTKEERESGLTNIVMGLSFITEGAIPFGAADPARAIPSFIAG 592
            GMVPPL + +AT LFK KFTK+ERE+G TN ++G SFITEGAIPF AADP R IPS I G
Sbjct: 505  GMVPPLGIALATTLFKKKFTKQEREAGKTNYILGASFITEGAIPFAAADPGRVIPSIIVG 564

Query: 593  SALTGALVGLAGIKLMAPHGGIFVI---ALTSNPILYLVFVVIGALVSGILFGALRKKA 648
            SA  G L  L  + L APHGG FVI    + +NP+LYLV ++ G++V+ +L G  +K A
Sbjct: 565  SAFAGGLTALFNVTLSAPHGGAFVIFIGNIVNNPLLYLVAIIAGSIVTALLLGFWKKDA 623
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 526/652 (80%), Positives = 581/652 (88%), Gaps = 6/652 (0%)

Query: 1    MKIQDLLKKEVMIMDLKATSKEAAIDEMITKLVDTGVVTNFAIFKDGIMKREAQTSTGLG 60
            MKIQDLL+K++MI+DL+A SKE AIDEMITKLV+  +V +F +FK  IM RE QTSTGLG
Sbjct: 1    MKIQDLLRKDIMILDLQAISKEVAIDEMITKLVEKDIVHDFDVFKKSIMTREEQTSTGLG 60

Query: 61   DGIAMPHSKNAAVKEATVLFAKSASGVDYEALDGQPTDLFFMIAAPDGANDTHLAALAEL 120
            DGIAMPHSKN  V +  VLFAKS  GVDY+ALDGQPTDLFFMIAAP GANDTHLAALAEL
Sbjct: 61   DGIAMPHSKNIVVDKPAVLFAKSNKGVDYKALDGQPTDLFFMIAAPQGANDTHLAALAEL 120

Query: 121  SKYLLKEGFADQLRQAKTPDDIIATFDSNSISQETVAPQTVQSTSKGSDYIVAVTACTTG 180
            S+YLLK+GFAD+LR A TP+ +IA FD  S ++E V   T      G D+IVAVTAC TG
Sbjct: 121  SQYLLKDGFADKLRAAATPEAVIAVFDEASTAKEEVVAPT-----SGQDFIVAVTACPTG 175

Query: 181  IAHTYMAEEALKKKAAEMGVGIKVETNGASGVGNKLTSSDIARAKGVIIAADKAVEMDRF 240
            IAHTYMAEEALKK+AAEMGV IKVETNGASGV N+LT+ DI RAKGVI+AADKAVEMDRF
Sbjct: 176  IAHTYMAEEALKKQAAEMGVAIKVETNGASGVANRLTAEDIQRAKGVIVAADKAVEMDRF 235

Query: 241  DGKPLVSRPVADGIKKSEDLINIILDNKAQTYHAKNQNDKQSGESDGKSGLGSAFYKHLM 300
            DGK  ++RPVADGIKKS++LI++IL+N+  TYHAKN    ++  S  K+ LG AFYKHLM
Sbjct: 236  DGKQFIARPVADGIKKSQELISLILNNEGNTYHAKN-GKSETAVSTEKTSLGGAFYKHLM 294

Query: 301  GGVSQMLPFVIGGGIMIAIAFLFDNILGVPKDQLSNLGSYHEIAALFKNIGGAAFAFMLP 360
            GGVSQMLPFVIGGGIMIA+AFL DN+LGVP DQL +LGSYHEIAA+F NIGGAAF+FMLP
Sbjct: 295  GGVSQMLPFVIGGGIMIALAFLLDNMLGVPNDQLGSLGSYHEIAAIFMNIGGAAFSFMLP 354
```

```
Query: 361 VLAGYIAYSIAEKPGLVAGFVAGSIASSGLAFGKVPFAEGGKATLALAGVPSGFLGALVG 420
           VLAGYIAYSIAEKPGLVAGFVAG+IAS+GLAFGKVPFA GG+ +L L GVPSGFLGALVG
Sbjct: 355 VLAGYIAYSIAEKPGLVAGFVAGAIASNGLAFGKVPFAAGGEVSLGLTGVPSGFLGALVG 414

Query: 421 GFLAGGVILLLRKLLSGLPKSLEGIKSILLYPLLGVLITGFLMLLVNIPMAAINTALNTF 480
           GFLAGGVIL LRKLL+GLP+SLEG+KSILLYPLLGVL+TGFLML VNIPMAAINTALN F
Sbjct: 415 GFLAGGVILALRKLLAGLPRSLEGVKSILLYPLLGVLVTGFLMLFVNIPMAAINTALNDF 474

Query: 481 LQGLSGSSAVLMGLLVGGMMAVDMGGPVNKAAYVFGTGTLAATVANGGSVVMAAVMAGGM 540
           LQGLSGSSAVLMGLLVGGMMAVDMGGPVNKAAYVFGTGTLAATVANGGSVVMAAVMAGGM
Sbjct: 475 LQGLSGSSAVLMGLLVGGMMAVDMGGPVNKAAYVFGTGTLAATVANGGSVVMAAVMAGGM 534

Query: 541 VPPLAVFVATLLFKDKFNNEERQSGLTNIVMGLSFITEGAIPFGAADPARAIPSFIVGSA 600
           VPPLAVFVATLLFKDKF  EER+SGLTNIVMGLSFITEGAIPFGAADPARAIPSFI GSA
Sbjct: 535 VPPLAVFVATLLFKDKFTKEERESGLTNIVMGLSFITEGAIPFGAADPARAIPSFIAGSA 594

Query: 601 LTGALVGLAGIKLMAPHGGIFVIALTSNPLLYILFILIGAVVSGVLFGLFRK 652
           LTGALVGLAGIKLMAPHGGIFVIALTSNP+LY++F++IGA+VSG+LFG  RK
Sbjct: 595 LTGALVGLAGIKLMAPHGGIFVIALTSNPILYLVFVVIGALVSGILFGALRK 646
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1954

A DNA sequence (GBSx2063) was identified in *S.agalactiae* <SEQ ID 6055> which encodes the amino acid sequence <SEQ ID 6056>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1532(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC24914 GB:AF012285 fructose-1-phosphate kinase [Bacillus subtilis]
 Identities = 146/303 (48%), Positives = 197/303 (64%)

Query: 1   MIYTVTLNPSIDFIVRLDTLLLGSVNRMTSDDKYVGGKGINVSRILKRLKIDNTATGFIG 60
           MIYTVTLNPS+D+IV ++   +G +NR + D KY GGKGINVSR+LKR  + + A GF+G
Sbjct: 1   MIYTVTLNPSVDYIVHVEDFTVGGLNRSSYDTKYPGGKGINVSRLLKRHHVASKALGFVG 60

Query: 61  GFTGHFVEDGLVLEGIKTDFVSVNEDTRINVKVKAKIETEINGGGPRITNEQLHRLEKLL 120
           GFTG +++  L  E ++T F  V  DTRINVK+K   ETEING GP I++E      +
Sbjct: 61  GFTGEYIKTFLREENLETAFSEVKGDTRINVKLKTGDETEINGQGPTISDEDFKAFLEQF 120

Query: 121 SRLTPEDTVVFAGSAPASLGNKVYNTLIPIAKKTGAEVVCDFEGQTLLDALAYQPLLVKP 180
             L   D VV AGS P+SL +  Y  +    K+  A VV D  G+ LL A   +P L+KP
Sbjct: 121 QSLQEGDIVVLAGSIPSSLPHDTYEKIAEACKQQNARVVLDISGEALLKATEMKPFLMKP 180

Query: 181 NNHELADIFGVELEGLPDIEKYAHKILDKGAKNVIVSMAGDGALLVTPEASYFAKPIKGE 240
           N+HEL ++FG  +  + +   Y  K++++GA++VIVSMAGDGALL T EA YFA   KG+
Sbjct: 181 NHHELGEMFGTAITSVEEAVPYGKKLVEQGAEHVIVSMAGDGALLFTNEAVYFANVPKGK 240

Query: 241 VKNSVGAGDSMVAGFTGEFVKSKNPVEALKWGVACGTATTFSDDLATAEFIQDIYNKVEV 300
           + NSVGAGDS+VAGF     K     EA + GV  G+AT FS++L T EF+Q +  +V+V
Sbjct: 241 LVNSVGAGDSVVAGFLAGISKQLPLEEAFRLGVTSGSATAFSEELGTEEFVQQLLPEVKV 300

Query: 301 EKL 303
            +L
Sbjct: 301 TRL 303
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6057> which encodes the amino acid sequence <SEQ ID 6058>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1738(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 222/302 (73%), Positives = 261/302 (85%)

Query: 1   MIYTVTLNPSIDFIVRLDTLLLGSVNRMTSDDKYVGGKGINVSRILKRLKIDNTATGFIG 60
           MIYTVTLNPSIDFIVR+D + LGSVNRM SDDK+ GGKGINVSRIL+RL I +TATGF+G
Sbjct: 1   MIYTVTLNPSIDFIVRIDQINLGSVNRMASDDKFAGGKGINVSRILQRLDIASTATGFLG 60

Query: 61  GFTGHFVEDGLVLEGIKTDFVSVNEDTRINVKVKAKIETEINGGGPRITNEQLHRLEKLL 120
           GFTG F+E+ L  EG+KTDFV  ++DTRINVK+K++ ETE+NG GP I+ EQL  L+  L
Sbjct: 61  GFTGRFIEESLSAEGVKTDFVKGDQDTRINVKIKSQEETELNGQGPIISQEQLEDLKTKL 120

Query: 121 SRLTPEDTVVFAGSAPASLGNKVYNTLIPIAKKTGAEVVCDFEGQTLLDALAYQPLLVKP 180
           S+LT EDTVVFAGSAPA+LGN VY  L+P+ +++GA+VVCDFEGQTL+DALAY PLLVKP
Sbjct: 121 SQLTAEDTVVFAGSAPANLGNAVYKELLPLVRQSGAQVVCDFEGQTLIDALAYNPLLVKP 180

Query: 181 NNHELADIFGVELEGLPDIEKYAHKILDKGAKNVIVSMAGDGALLVTPEASYFAKPIKGE 240
           NNHEL  IFG  L  L D+E YA ++L+ GA+NVI+SMAGDGALLVT EA+YFAKPIKGE
Sbjct: 181 NNHELEAIFGTILTSLDDVETYARRLLEMGAQNVIISMAGDGALLVTKEATYFAKPIKGE 240

Query: 241 VKNSVGAGDSMVAGFTGEFVKSKNPVEALKWGVACGTATTFSDDLATAEFIQDIYNKVEV 300
           VKNSVGAGDSMVAGFTGEF+KS+NP+EALKWGVACGTAT FSDDLAT  FI++ Y+KVEV
Sbjct: 241 VKNSVGAGDSMVAGFTGEFMKSQNPIEALKWGVACGTATAFSDDLATIAFIKETYHKVEV 300

Query: 301 EK 302
           EK
Sbjct: 301 EK 302
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1955

A DNA sequence (GBSx2064) was identified in *S.agalactiae* <SEQ ID 6059> which encodes the amino acid sequence <SEQ ID 6060>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2769(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9961> which encodes amino acid sequence <SEQ ID 9962> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC24913 GB:AF012285 FruR [Bacillus subtilis]
 Identities = 97/247 (39%), Positives = 148/247 (59%), Gaps = 4/247 (1%)

Query: 23  MLKSKRKEIILSRLEQNKSVTLDELTSILETSESTVRRDLDELESAGFLKRVHGGAELPY 82
```

```
           ML  +R ++I+ ++E++  V + EL ++   SEST+RRDL  LE  GFLKRVHGGA
Sbjct: 1   MLTPERHQLIIDQIEKHDVVKIQELINLTNASESTIRRDLSTLEERGFLKRVHGGAAKLS 60

Query: 83  SLGQELSNQEKAIKNVQKKLDIARQTAKLIAKQDVIFIDAGTTTELLIDFLPH-EQLTVV 141
            +  E    EK+ KN+  KL IA + A L+ + D I++DAGTTT  +IDF+   + + VV
Sbjct: 61  DIRLEPDMLEKSSKNLHDKLKIAEKAASLLEEGDCIYLDAGTTTLHMIDFMDKTKDIVVV 120

Query: 142 TNSIHHAAKLVDRGIKTIIIGGAVKHSTDASIGQVAINQIRQITVDKAFLGMNGID-EVY 200
           TN + H   L+ + I   ++GG VKH T A IG  ++  + Q   DK+FLG NG+  E
Sbjct: 121 TNGVMHIDALIRKEISFYLLGGYVKHRTGAIIGGASLVAMDQYRFDKSFLGTNGVHTEAG 180

Query: 201 LTTPDLEEAAIKEAIINNSQQTFILMDSSKIGQVTFAKVKEINDINLVTNKTDSELMTII 260
            TTPD +EA +K+  I  ++  ++L D SK G+++F+    I D  ++T  TD+E +T
Sbjct: 181 FTTPDPDEALLKQKAIKQAKHAYVLADPSKFGEISFSAFAGIGDATIIT--TDAEELTFD 238

Query: 261 KEKMKVI 267
             + K +
Sbjct: 239 NYQEKTV 245
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6061> which encodes the amino acid sequence <SEQ ID 6062>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2604(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 135/237 (56%), Positives = 184/237 (76%)

Query: 33  LSRLEQNKSVTLDELTSILETSESTVRRDLDELESAGFLKRVHGGAELPYSLGQELSNQE 92
           ++++ +   V+L++L  +L +SEST+RRDL ELE  G L RVHGGAEL +SL +ELSNQE
Sbjct: 1   MAKITEENYVSLEDLMQLLNSSESTIRRDLGELEQEGRLHRVHGGAELFHSLQEELSNQE 60

Query: 93  KAIKNVQKKLDIARQTAKLIAKQDVIFIDAGTTTELLIDFLPHEQLTVVTNSIHHAAKLV 152
           K++KN   K  IA++ ++LI   DVIFIDAGTTTE L+ FL  + LTVVTNSIHHAA+LV
Sbjct: 61  KSVKNSHIKKAIAQRASQLIYDNDVIFIDAGTTTEFLLPFLQAKNLTVVTNSIHHAARLV 120

Query: 153 DRGIKTIIIGGAVKHSTDASIGQVAINQIRQITVDKAFLGMNGIDEVYLTTPDLEEAAIK 212
           +  I+TII+GG VK +TDASIG VA+ QIRQ+  DKAFLGMNG+D+ YLTTPD+EEA IK
Sbjct: 121 ELSIETIIVGGYVKQTTDASIGNVALEQIRQMNFDKAFLGMNGVDDSYLTTPDMEEAVIK 180

Query: 213 EAIINNSQQTFILMDSSKIGQVTFAKVKEINDINLVTNKTDSELMTIIKEKMKVIQV 269
           +A+++N++  +IL+D +KIGQV+F KV  IND+ ++T    + ++  IKEK KVI++
Sbjct: 181 KAVLSNAKLAYILVDGTKIGQVSFVKVAPINDVTIITLGGSASILKQIKEKAKVIEL 237
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1956

A DNA sequence (GBSx2065) was identified in *S.agalactiae* <SEQ ID 6063> which encodes the amino acid sequence <SEQ ID 6064>. This protein is predicted to be beta-lactam resistance factor. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5777(Affirmative) < succ>
```

```
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB89121 GB:AJ277485 beta-lactam resistance factor
          [Streptococcus pneumoniae]
 Identities = 215/410 (52%), Positives = 283/410 (68%)

Query: 1   MTLRELTIEEFKEHSGNYDSQSFLQTPEMAKLLEKRGYDVRYLGYQVENKLEIISLSYIM 60
           M L  LT EEF+ +S    S+SF+Q+ +M  LLEKRG  + YL  + E ++++ +L Y +
Sbjct: 1   MALTTLTKEEFQTYSDQVSSRSFMQSVQMGDLLEKRGARIVYLALKQEGEIQVAALVYSL 60

Query: 61  PVTGGFQMKIDSGPVHSNSKYLKQFYKALQGYAKSNGVLELIVEPYDDYQLFTSSGVPSN 120
           P+ GG  M+++SGP+++    L  FY  L+ YAK NGVLEL+V+PY+ YQ F S G P +
Sbjct: 61  PMLGGLHMELNSGPIYTQQDALPVFYAELKEYAKQNGVLELLVKPYETYQTFDSQGNPID 120

Query: 121 QGNDNLIEDFTSSGYHHDGLTTGFTGKYLSWHYVKNLEGVTSETLLSSFSKTGRALVKKA 180
               ++I+D T  GY  DGLTTG+ G    W Y K+L  +T ++LL SFSK G+ LVKKA
Sbjct: 121 AEKKSIIQDLTDLGYQFDGLTTGYPGGEPDWLYYKDLTELTEKSLLKSFSKKGKPLVKKA 180

Query: 181 MSFGIKVRVLKRDELHLFKEITTSTSNRRDYMDKSLDYYQDFYDSFEGKAEFVIATLNFR 240
            +FGI+++ LKR+EL +FK IT  TS RR+Y DKSL+YY+ FYD+F  +AEF+IA+LNF
Sbjct: 181 ETFGIRLKKLKREELSIFKNITKETSERREYSDKSLEYYEHFYDTFGEQAEFLIASLNFS 240

Query: 241 EYDHNLQIKAEALENKLKLLDERFRENADSPKYHRQRSEIINQLASFETRRQEVQSFIQK 300
           +Y   LQ +   LE  L  L     +N  S K   Q  E  +Q  +FE R+ E +  I+K
Sbjct: 241 DYMSKLQGEQSKLEENLDKLRLDLSKNPHSEKKQNQLREYSSQFETFEVRKAEARDLIEK 300

Query: 301 YDNQDVVLAGSLFVYSLKETVYFFSGSYTEFNKFYAPAVLQEYVMQEALKRGSTFYNLLG 360
           Y  +D+VLAGSLFVY  +ET Y FSGSYTEFNKFYAPA+LQ+YVM E++KRG   YN LG
Sbjct: 301 YGEEDIVLAGSLFVYMPQETTYLFSGSYTEFNKFYAPALLQKYVMLESIKRGIPKYNFLG 360

Query: 361 IQGTFDGSDSILRFKQNFNGCIIRKMGTFNYYPSPFKYKGIQLLKKVLKR 410
           IQG FDGSD +LRFKQNFNG I+RK GTF Y+PSP KYK IQLLKK++ R
Sbjct: 361 IQGIFDGSDGVLRFKQNFNGYIVRKAGTFRYHPSPLKYKAIQLLKKIVGR 410
```

There is also homology to SEQ ID 5460.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1957

A DNA sequence (GBSx2066) was identified in *S.agalactiae* <SEQ ID 6065> which encodes the amino acid sequence <SEQ ID 6066>. This protein is predicted to be cell wall protein, 40 kDa (sr 5' region). Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.45   Transmembrane   25 -  41 (  23 -  42)

----- Final Results -----
               bacterial membrane --- Certainty=0.2381(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9963> which encodes amino acid sequence <SEQ ID 9964> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
!GB:AF278686 choline binding protein D; CbpD [Strept...
!GB:AF278686 choline binding protein D; CbpD [Strept...
```

```
>GP:AAF87768 GB:AF278686 choline binding protein D; CbpD
           [Streptococcus pneumoniae]
 Identities = 63/230 (27%), Positives = 108/230 (46%), Gaps = 34/230 (14%)

Query: 324 WTEQGGQDDIKWYTAVTTGDG------NYKVAVSFADHKNEKGLYNIHLYYQEASGTLVG 377
            W+  G    + W + V  GD       NY    S+     +       +++++  G  VG
Sbjct: 123 WSTAGTYGHVAWVSNVM-GDQIEIEEYNYGYTESYNKRVIKANTMTGFIHFKDLDGGSVG 181

Query: 378 VTGTKVTVAGTNSSQEPIENGLAKTGVYNIIGSTEVKNEAKISSQTQFTLEKGDKINYDQ 437
             + +  +  GT+  +                  + +K E   S         G+K++YDQ
Sbjct: 182 NSQSSTSTGGTHYFKT----------------KSAIKTEPLASGTVIDYYYPGEKVHYDQ 225

Query: 438 VLTADGYQWISYKSYSGVRRYIPVKKLTTSSEKAKDEATKPTSYPNLPKTG-TYTFTKTV 496
            +L  DGY+W+SY +Y+G  RY+ ++ +  +          P     L  TG T+ F
Sbjct: 226 ILEKDGYKWLSYTAYNGSYRYVQLEAVNKN----------PLGNSVLSSTGGTHYFKTKS 275

Query: 497 DVKSQPKVSSPVEFNFQKGEKIHYDQVLVVDGHQWISYKSYSGIRRYIEI 546
             +K++P VS+ V   +  GEK+HYDQ+L  DG++W+SY +Y+G RRYI++
Sbjct: 276 AIKTEPLVSATVIDYYYPGEKVHYDQILEKDGYKWLSYTAYNGSRRYIQL 325
 Identities = 49/161 (30%), Positives = 85/161 (52%), Gaps = 14/161 (8%)

Query: 116 GNYVYSKETEVKNTPSKSAPVAFYAKKGDKVFYDQVFNKDNVKWISYKSFCGVRRYAAIE 175
            G + +  ++ +K  P  S  V  Y   G+KV YDQ+  KD  KW+SY ++ G  RY  +E
Sbjct: 191 GTHYFKTKSAIKTEPLASGTVIDYYYPGEKVHYDQILEKDGYKWLSYTAYNGSYRYVQLE 250

Query: 176 SLDPSGGSETKAPTPVTNSGSNNQEKIATQGNYTFSHKVEVKNEAKVASPTQFTLDKGDR 235
            +++ +         P+ NS  +     +T G + F  K  +K E  V++        G++
Sbjct: 251 AVNKN---------PLGNSVLS-----STGGTHYFKTKSAIKTEPLVSATVIDYYYPGEK 296

Query: 236 IFYDQILTIEGNQWLSYKSFNGVRRFVLLGKASSVEKTEDK 276
            + YDQIL  +G +WLSY ++NG RR++ L   +S +  +++
Sbjct: 297 VHYDQILEKDGYKWLSYTAYNGSRRYIQLEGVTSSQNYQNQ 337
 Identities = 52/192 (27%), Positives = 90/192 (46%), Gaps = 13/192 (6%)

Query: 295 ISNETTTGFDILITNIKDDNGIAAVKVPVWTEQGGQDDIKWYTAVTTGDGNYKVAVSFAD 354
            I   T TGF     + KD +G +       T  GG    K  +A+ T        + +
Sbjct: 161 IKANTMTGF----IHFKDLDGGSVGNSQSSTSTGGTHYFKTKSAIKTEPLASGTVIDYY- 215

Query: 355 HKNEKGLYNIHLY---YQEASGTLVGVTGTKVTVAGTNSSQEPIENGLAKT--GVYNIIG 409
            +  EK  Y+  L    Y+  S T    +   V +   N +  P+ N +  +  G +
Sbjct: 216 YPGEKVHYDQILEKDGYKWLSYTAYNGSYRYVQLEAVNKN--PLGNSVLSSTGGTHYFKT 273

Query: 410 STEVKNEAKISSQTQFTLEKGDKINYDQVLTADGYQWISYKSYSGVRRYIPVKKLTTSSE 469
             + +K E  +S+        G+K++YDQ+L  DGY+W+SY +Y+G RRYI ++ + TSS+
Sbjct: 274 KSAIKTEPLVSATVIDYYYPGEKVHYDQILEKDGYKWLSYTAYNGSRRYIQLEGV-TSSQ 332

Query: 470 KAKDEATKPTSY 481
              ++++   +SY
Sbjct: 333 NYQNQSGNISSY 344
 Identities = 33/113 (29%), Positives = 56/113 (49%), Gaps = 2/113 (1%)

Query: 91  NTATKDITTPLVETKPMVEKTLPEQGNYVYSK-ETEVKNTPSKSAPVAFYAKKGDKVFYD 149
            N + + +    V   P+    L   G   Y K ++ +K  P  SA V  Y   G+KV YD
Sbjct: 241 NGSYRYVQLEAVNKNPLGNSVLSSTGGTHYFKTKSAIKTEPLVSATVIDYYYPGEKVHYD 300

Query: 150 QVFNKDNVKWISYKSFCGVRRYAAIESLDPSGGSETKAPTPVTNSGSNNQEKI 202
            Q+  KD  KW+SY ++ G RRY  +E +  S   + ++   +++ GS++   +
Sbjct: 301 QILEKDGYKWLSYTAYNGSRRYIQLEGVTSSQNYQNQSGN-ISSYGSHSSSTV 352
```

A related GBS gene <SEQ ID 8937> and protein <SEQ ID 8938> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:     -6.74
GvH: Signal Score (-7.5): 1.26
     Possible site: 42
```

```
>>> Seems to have no N-terminal signal sequence
ALOM program  count: 1 value:  -3.45 threshold:  0.0
   INTEGRAL    Likelihood = -3.45   Transmembrane   22 -  39 (  23 -  42)
   PERIPHERAL  Likelihood =  6.26     371
 modified ALOM score:   1.19

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.2381(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear)
```

The protein has homology with the following sequences in the databases:

```
41.2/57.9% over 283aa
                                                                  Streptococ
cus mutans
  EGAD|33594| cell wall protein, 40 kDa (sr 5' region) Insert characterized
  PIR|A60328|A60328 40K cell wall protein precursor (sr 5' region) - (strain OMZ
175, serotype f) Insert characterized

ORF02145(301 - 1803 of 2238)
EGAD|33594|34911(30 - 313 of 335) cell wall protein, 40 kDa (sr 5' region) {Stre
ptococcus mutans}PIR|A60328|A60328 40K cell wall protein precursor (sr 5' region
) - Streptococcus mutans (strain OMZ175, serotype f)
%Match = 8.0
%Identity = 41.1  %Similarity = 57.9
Matches = 81  Mismatches = 79  Conservative Sub.s = 33

156       186       216       246       276       306       336       366
*YA****FCYTKNNKSWVFFSRSIYSIKYYICITNISKIC*HVTKRIL***CK*IRK*VFMMKKGQVNDTKQSYSLRKYK
                                   :      :    :        : | :||:  :|   | :|:|||
                             MNQKIVVISSFYMLGAHSFSKAVYHNDRSVKLMKRIDINHQAQRFSIRKYA
                                      10        20        30        40        50

396       426       456       486       516       546       576       606
FGLASVILGSFIMVTSPVFADQTTSVQVNNQTGTSVDANNSSNETSASSVITSNNDSVQASDKVVNSQNTATKDITTPLV
|| |||::|    : :   : |    |:                 | ::|:    :||  |  | :   | |    || |
FGAASVLIGCVFFLGTQNVSAQEQGTQL---------------PASENAVVNVAENSVAISQAVADKAATQTTLTETPQV
                  70                       80        90       100       110

654       684       714
ETKPMVEK--------------------~~~~--------------------TLPEQGNYVYSKETEVKNTPSKSAPVAF
| :    |                                            ::| |||||: : | |||  | |:|
EVEEKESKVNAPALNVDDKGAKSKEDVN~~~~AEQNEKAVRENLMCRQAKAVSIPSQGNYVFQETTPVKNAASMSSP---
            130       140          200       210       220       230       240

744                1533      1563      1593      1623      1653      1683
YAKKGDKVFYDQVFNKD~~~~GVYNIIGSTEVKNEAKISSQTQFTLEKGDKINYDQVLTADGYQWISYKSYSGVRRYIPV
                                         ||| ::||||: || || |||:||||| ||||:||| |:
-----------------~~~~--------------------TQFNFDKGDKVFYDNVLEADGHQWISYVSYSGIRRYAPI
                                                 250       260       270

1713      1743      1773      1803      1833      1863      1893      1923
KKLTTSSEKAKDEATKPTSYPNLPKTGTYTFTKTVDVKSQPKVSSPVEFNFQKGEKIHYDQVLVVDGHQWISYKSYSGIR
      :   ::   |     |||  ||| |||   :| :          |         :
------AVTIEELKQKEIVQQNLPAQGTYHFTKQQSLKMKLNCLVRPNSRFTTEITFFMIRF
               290       300       310       320       330
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6067> which encodes the amino acid sequence <SEQ ID 6068>. Analysis of this protein sequence reveals the following:

```
          Possible site: 23

  >>> Seems to have a cleavable N-term signal seq.
```

```
----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF87768 GB:AF278686 choline binding protein D; CbpD
           [Streptococcus pneumoniae]
 Identities = 93/217 (42%), Positives = 136/217 (61%), Gaps = 18/217 (8%)

Query: 42  GDNYPSKWKKGNG-IDSWNMYIRQCTSFAAFRLSSANGFQLPKGYGNACTWGHIAKNQGY 100
           GD+YP+ +K G+  ID W MY RQCTSF AFRLS+ NGF++P  YGNA  WGH A+ +GY
Sbjct: 51  GDDYPAYYKNGSQEIDQWRMYSRQCTSFVAFRLSNVNGFEIPAAYGNANEWGHRARREGY 110

Query: 101 PVNKTPSIGAIAWFDKNAYQSNAAYGHVAWVADIRGDTVTIEEYNYNAGQGPERYHKRQI 160
            V+ TP+IG+I W       +   YGHVAWV+++ GD + IEEYNY      E Y+KR I
Sbjct: 111 RVDNTPTIGSITW------STAGTYGHVAWVSNVMGDQIEIEEYNYGY---TESYNKRVI 161

Query: 161 PKSQVSGYIHFKDLSSQTSHSYPRQLKHISQASFDPSGTYHFTTRLPVKGQTSIDSPDLA 220
             + ++G+IHFKDL   +  +        SQ+S    GT++F T+  +K +       +
Sbjct: 162 KANTMTGFIHFKDLDGGSVGN--------SQSSTSTGGTHYFKTKSAIKTEPLASGTVID 213

Query: 221 YYEAGQSVYYDKVVTAGGYTWLSYLSFSGNRRYIPIK 257
           YY  G+ V+YD+++   GY WLSY +++G+ RY+ ++
Sbjct: 214 YYYPGEKVHYDQILEKDGYKWLSYTAYNGSYRYVQLE 250
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 34/94 (36%), Positives = 52/94 (55%)

Query: 453 SGVRRYIPVKKLTTSSEKAKDEATKPTSYPNLPKTGTYTFTKTVDVKSQPKVSSPVEFNF 512
           S V  YI  K L++ +  +     K  S  +   +GTY FT  + VK Q  + SP    +
Sbjct: 163 SQVSGYIHFKDLSSQTSHSYPRQLKHISQASFDPSGTYHFTTRLPVKGQTSIDSPDLAYY 222

Query: 513 QKGEKIHYDQVLVVDGHQWISYKSYSGIRRYIEI 546
           + G+ ++YD+V+   G+ W+SY S+SG RRYI I
Sbjct: 223 EAGQSVYYDKVVTAGGYTWLSYLSFSGNRRYIPI 256
 Identities = 30/78 (38%), Positives = 45/78 (57%), Gaps = 2/78 (2%)

Query: 402 TGVYNIIGSTEVKNEAKISSQTQFTLEKGDKINYDQVLTADGYQWISYKSYSGVRRYIPV 461
           +G Y+      VK +  I S      E G  + YD+V+TA GY W+SY S+SG RRYIP+
Sbjct: 197 SGTYHFTTRLPVKGQTSIDSPDLAYYEAGQSVYYDKVVTAGGYTWLSYLSFSGNRRYIPI 256

Query: 462 KKLTTSSEKAKDEATKPT 479
           K+   +    +++ TKP+
Sbjct: 257 KE--PAQSVVQNDNTKPS 272
 Identities = 27/94 (28%), Positives = 47/94 (49%)

Query: 198 NQEKIATQGNYTFSHKVEVKNEAKVASPTQFTLDKGDRIFYDQILTIEGNQWLSYKSFNG 257
           +Q      G Y F+ ++ VK +  + SP     + G  ++YD+++T  G  WLSY SF+G
Sbjct: 190 SQASFDPSGTYHFTTRLPVKGQTSIDSPDLAYYEAGQSVYYDKVVTAGGYTWLSYLSFSG 249

Query: 258 VRRFVLLGKASSVEKTEDKEKVSPQPQARITKTG 291
            RR++ + + +      D  K S +    +T  G
Sbjct: 250 NRRYIPIKEPAQSVVQNDNTKPSIKVGDTVTFPG 283
 Identities = 23/73 (31%), Positives = 35/73 (47%)

Query: 103 ETKPMVEKTLPEQGNYVYSKETEVKNTPSKSAPVAFYAKKGDKVFYDQVFNKDNVKWISY 162
           + K + + +    G Y ++    VK   S  +P   Y + G  V+YD+V       W+SY
Sbjct: 185 QLKHISQASFDPSGTYHFTTRLPVKGQTSIDSPDLAYYEAGQSVYYDKVVTAGGYTWLSY 244

Query: 163 KSFCGVRRYAAIE 175
            SF G RRY  I+
Sbjct: 245 LSFSGNRRYIPIK 257
```

SEQ ID 8938 (GBS91) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 18 (lane 7; MW 63kDa).

The GBS91-His fusion product was purified (Figure 195, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 283), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1958

A DNA sequence (GBSx2067) was identified in *S.agalactiae* <SEQ ID 6069> which encodes the amino acid sequence <SEQ ID 6070>. This protein is predicted to be thiamine biosynthesis protein. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0984(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB49673 GB:AJ248285 PROBABLE 2-DEHYDROPANTOATE 2-REDUCTASE (EC
           1.1.1.169) [Pyrococcus abyssi]
 Identities = 85/301 (28%), Positives = 150/301 (49%), Gaps = 7/301 (2%)

Query: 1   MLVYIAGSGAMGCRFGYQISKTNHDVILLDNWADHIMAIKENGLKVTGDTEDLVKLPIMK 60
           M +YI G+GA+G  FG  ++    DV+L+     H+ AI E GLK+ G  +  VK+
Sbjct: 1   MKIYILGAGAIGSLFGGLLANAGEDVLLIGR-DPHVSAINEKGLKIVGIKDLNVKVEATT 59

Query: 61  PTDATEEADLIILFTKAMQLPNMLQDIKKIIGKKTKVLCLLNGLGHEDVIRQYIPEHNIL 120
                E+ DLI+L TK+      L+  + I+ K + VL + NG+G+ED I ++      +
Sbjct: 60  RVPE-EKPDLIVLATKSYSTIEALKSARHIV-KGSWVLSIQNGIGNEDKIIEF--GGKAI 115

Query: 121 MGVTVWTAGLKGPGHAHLEGVGSVNLQSIDPNNQEAGHRVTELLNEAKLQATYDENVLPN 180
            G+T   A ++ PG     G G   +       ++   +V ++ N A ++    EN++
Sbjct: 116 GGITTNGAMVEAPGVIKWTGKGVTIIGLYPQGKEKFIEKVADVFNSADIETHVSENIISW 175

Query: 181 IWRKACVNGTMNSTCALLDCTIGQLFASEDGVNMVHEIIHEFVTVGKAEGVELDEEEITK 240
           IW KA VN  +N    LL+     +  ++  ++M  E++ E   V    G+E D   +
Sbjct: 176 IWAKAIVNSAINPIGTLLEVKNKVIRENDFLLSMAMEVVKEGCRVALQNGIEFDVPPMDL 235

Query: 241 YVMDTSVKAAHHYPSMHQDLVQNQRLTEIDFLNGAVNKKGENLGIDTPYCRLITQLIHTKE 301
           +   T  +   +Y SM QD+ + ++ TE+D++NG + +  + + ++ P   L+  LI  KE
Sbjct: 236 F-FQTLEQTRENYNSMLQDIWRGKK-TEVDYINGKIVEYAKAVNLEAPMNLLLWGLIKGKE 294
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6071> which encodes the amino acid sequence <SEQ ID 6072>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1392(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 262/307 (85%), Positives = 288/307 (93%)

Query: 1   MLVYIAGSGAMGCRFGYQISKTNHDVILLDNWADHIMAIKENGLKVTGDTEDLVKLPIMK 60
            MLVYIAGSGAMGCRFGYQISKTN+DVILLDNW DHI AIKENGL VTGD E+ VKLPIMK
Sbjct: 1   MLVYIAGSGAMGCRFGYQISKTNNDVILLDNWEDHINAIKENGLVVTGDVEETVKLPIMK 60

Query: 61  PTDATEEADLIILFTKAMQLPNMLQDIKKIIGKKTKVLCLLNGLGHEDVIRQYIPEHNIL 120
            PT+AT+EADLIILFTKAMQLP MLQDIK IIGK+TKVLCLLNGLGHEDVIRQYIPEHNIL
Sbjct: 61  PTEATQEADLIILFTKAMQLPQMLQDIKGIIGKETKVLCLLNGLGHEDVIRQYIPEHNIL 120

Query: 121 MGVTVWTAGLKGPGHAHLEGVGSVNLQSIDPNNQEAGHRVTELLNEAKLQATYDENVLPN 180
            MGVTVWTAGL+GPG AHL+GVG++NLQS+DP+NQEAGH+V +LLNEA L ATYDENV+PN
Sbjct: 121 MGVTVWTAGLEGPGRAHLQGVGALNLQSMDPSNQEAGHQVADLLNEANLNATYDENVVPN 180

Query: 181 IWRKACVNGTMNSTCALLDCTIGQLFASEDGVNMVHEIIHEFVTVGKAEGVELDEEEITK 240
            IWRKACVNGTMNSTCALLDCTIG+LFASEDG+ MV EIIHEFV VG+AEGVEL+EEEIT+
Sbjct: 181 IWRKACVNGTMNSTCALLDCTIGELFASEDGLKMVKEIIHEFVIVGQAEGVELNEEEITQ 240

Query: 241 YVMDTSVKAAHHYPSMHQDLVQNQRLTEIDFLNGAVNKKGENLGIDTPYCRLITQLIHTK 300
            YVMDTSVKAAHHYPSMHQDLVQN RLTEIDF+NGAVN KGE LGI+TPYCR+IT+L+H K
Sbjct: 241 YVMDTSVKAAHHYPSMHQDLVQNHRLTEIDFINGAVNTKGEKLGINTPYCRMITELVHAK 300

Query: 301 ENVLSIK 307
            E VL+I+
Sbjct: 301 EAVLNIQ 307
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1959

A DNA sequence (GBSx2068) was identified in *S.agalactiae* <SEQ ID 6073> which encodes the amino acid sequence <SEQ ID 6074>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.03   Transmembrane   61 -  77 (  61 -  78)
   INTEGRAL    Likelihood = -1.33   Transmembrane   80 -  96 (  79 -  96)

----- Final Results -----
               bacterial membrane --- Certainty=0.2211(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1960

A DNA sequence (GBSx2069) was identified in *S.agalactiae* <SEQ ID 6075> which encodes the amino acid sequence <SEQ ID 6076>. This protein is predicted to be regulatory protein (pfoS/R). Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.82   Transmembrane  317 - 333 ( 304 - 335)
   INTEGRAL    Likelihood = -7.64   Transmembrane  187 - 203 ( 183 - 217)
   INTEGRAL    Likelihood = -5.26   Transmembrane   24 -  40 (  18 -  44)
   INTEGRAL    Likelihood = -5.04   Transmembrane  143 - 159 ( 139 - 161)
```

```
  INTEGRAL    Likelihood = -2.34   Transmembrane  116 - 132 ( 115 - 136)
  INTEGRAL    Likelihood = -2.13   Transmembrane   55 -  71 (  55 -  71)
  INTEGRAL    Likelihood = -0.96   Transmembrane  268 - 284 ( 268 - 284)

----- Final Results -----
               bacterial membrane --- Certainty=0.4927(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC65034 GB:AE001189 regulatory protein (pfoS/R) [Treponema
          pallidum]
 Identities = 138/358 (38%), Positives = 220/358 (60%), Gaps = 18/358 (5%)

Query: 2   TNTVTPKETAGSFINKVLGGTATAIVVALIPNAILATFLKPFLSYG-LAAEFLHIVQVFQ 60
            T +++P++    F+ K+L G++  IV+ L+P AI     +       L A   H+V   Q
Sbjct: 3   TQSLSPRQ----FMMKILNGSSAGIVIGLVPPAIAGELFRALAPLSPLFAALYHVVLPIQ 58

Query: 61  FFTPIMAGFLIGQQFKFTPMQQLAVGGAAYIGSGAWAYTEVIQKGVATGSFQLRGIGDLI 120
            F  P + G L+G QF  +  +   +   + I SG           +  G++ + GIGD+I
Sbjct: 59  FSVPALIGTLVGLQFHCSAPEVATLAFVSVIASG--------NVTLQNGAWLITGIGDVI 110

Query: 121 NMMLTAALAVLAVKWFGNKFGSLTIILLPIIIGTGVGYLGWKLLPYVSYVTTLIGQGINS 180
            N+ML +ALA++ V+    K GSLTII LP+I+    G +G   LPYV  +T  +G+ I +
Sbjct: 111 NVMLISALAIILVRALRGKLGSLTIIALPVIVAVVAGGVGSFSLPYVKMITLFVGRVIAT 170

Query: 181 FTTLQPIAMSILIAMAFSMLIVSPISTVAIGLAIGLNGMSASAASMGVASTTAVLVWATM 240
            F  LQP+ MSIL++M+FS++I+SP+S+VA+G+A+GL G+++ AA++GV+S    L+  TM
Sbjct: 171 FIALQPLLMSILLSMSFSLIIISPVSSVAVGIAVGLTGLASGAANIGVSSCAMTLIVGTM 230

Query: 241 KANKSGVPIAIALGAMKMMMPNFLKHPVMAIPMLMTATVSSLTVPLFKLVGTPASSGFGL 300
            + NK GVP+A+  GAMKM+MPN++++P++ IP+L+   V  +   LF L GTPAS+GFG
Sbjct: 231 RVNKIGVPLAMFAGAMKMLMPNWIRYPILNIPLLLNGLVCGVLAWLFNLQGTPASAGFGF 290

Query: 301 VGAVGPIASFE--AGASML---IVILSWLVIPFAVGFVSHKICKDILKLYKDDIFVFE 353
            +G VGPI ++   A   M+   I+ L + V+ F   ++   I  D LKLY+ ++F+ E
Sbjct: 291 IGLVGPINAYRLMAYTPMVRAGILFLVYFVLSFLAAYLIDFILVDRLKLYRRELFIPE 348
```

There is also homology to SEQ ID 1280.

A related GBS gene <SEQ ID 8939> and protein <SEQ ID 8940> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:     -7.24
GvH: Signal Score (-7.5): -2.94
     Possible site: 49
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 7 value:  -9.82 threshold:  0.0
   INTEGRAL    Likelihood = -9.82   Transmembrane  317 - 333 ( 304 - 335)
   INTEGRAL    Likelihood = -7.64   Transmembrane  187 - 203 ( 183 - 217)
   INTEGRAL    Likelihood = -6.37   Transmembrane  143 - 159 ( 136 - 161)
   INTEGRAL    Likelihood = -5.26   Transmembrane   24 -  40 (  18 -  44)
   INTEGRAL    Likelihood = -2.34   Transmembrane  116 - 132 ( 115 - 136)
   INTEGRAL    Likelihood = -2.13   Transmembrane   55 -  71 (  55 -  71)
   INTEGRAL    Likelihood = -0.96   Transmembrane  268 - 284 ( 268 - 284)
   PERIPHERAL  Likelihood =  0.69     205
 modified ALOM score:   2.46

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4927(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF02147(337 - 1359 of 1668)
EGAD|138195|TP0038(10 - 348 of 350) regulatory protein {Treponema pallidum} OMNI|TP0038
regulatory protein (pfoS/R) GP|3322295|gb|AAC65034.1||AE001189 regulatory protein (pfoS/R)
{Treponema pallidum} PIR|E71373|E71373 probable regulatory protein (pfoS/R) - syphilis
spirochete
%Match = 21.6
%Identity = 40.1  %Similarity = 65.6
Matches = 135  Mismatches = 112  Conservative Sub.s = 86

87        117       147       177       207       237       267       297
LQQDMGKHQSL*TKLSIIFILIEITV*SIQHH**NNYK*N**VYKKGLYILLKK*QSFLFIL*YN*LCRYE*Y*INEARY

327       357       387       417       444       474       504       534
FMTNTVTPKETAGSFINKVLGGTATAIVVALIPNAILATFLKPFLSYG-LAAEFLHIVQVFQFFTPIMAGFLIGQQFKFT
              |: |:| |::  ||: |:| ||   ::: :     | | : |:|   ||  | : | |:| ||  :
    MHTQSLSPRQFMMKILNGSSAGIVIGLVPPAIAGELFRALAPLSPLFAALYHVVLPIQFSVPALIGTLVGLQFHCS
            10        20        30        40        50        60        70

564       594       624       654       684       714       744       774
PMQQLAVGGAAYIGSGAWAYTEVIQKGVATGSFQLRGIGDLINMMLTAALAVLAVKWFGNKFGSLXIILLPIIIGTGVGY
  :   :   : | ||           :  |:: : ||||:||:|| :|||:: |: :  |:||| || ||:|:    |
APEVATLAFVSVIASG--------NVTLQNGAWLITGIGDVINVMLISALAIILVRALRGKLGSLTIIALPVIVAVVAGG
            90               100       110       120       130       140

804       834       864       894       924       954       984       1014
LGWKLLPYVSYVTTLIGQGINSFTTLQPIAMSILIAMAFSMLIVSPISTVAIGLAIGLNGMSASAASMGVASTTAVLVWA
:|   ||||  :| ::|: | :|  |||: ||||::|:||::|:||:|:||:|:|:|| |::: ||::||:|    |:
VGSFSLPYVKMITLFVGRVIATFIALQPLLMSILLSMSFSLIIISPVSSVAVGIAVGLTGLASGAANIGVSSCAMTLIVG
         160       170       180       190       200       210       220

1044      1074      1104      1134      1164      1194      1224      1248
TMKANKSGVPIAIALGAMKMMMPNFLKHPVMAIPMLMTATVSSLTVPLFKLVGTPASSGFGLVGAVGPIASFE--AGASM
||: || |||:|:  |||||:|||::::|:: ||:|:   |  :   || |||||:|||::| |||| ::   |   |
TMRVNKIGVPLAMFAGAMKMLMPNWIRYPILNIPLLLNGLVCGVLAWLFNLQGTPASAGFGFIGLVGPINAYRLMAYTPM
         240       250       260       270       280       290       300

1269      1299      1329      1359      1389      1419      1449      1479
L---IVILSWLVIPFAVGFVSHKICKDILKLYKDDIFVFEGQN*FGGCMLVYIAGSGAMGCRFGYQISKTNHDVILLDNW
:   |: | ::|: |   ::   |  | ||||: ::|: |
VRAGILFLVYFVLSFLAAYLIDFILVDRLKLYRRELFIPEQG
         320       330       340       350
```

There is also homology to SEQ ID 1276

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1961

A DNA sequence (GBSx2070) was identified in *S.agalactiae* <SEQ ID 6077> which encodes the amino acid sequence <SEQ ID 6078>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07127 GB:AP001518 thioredoxin reductase [Bacillus halodurans]
 Identities = 163/325 (50%), Positives = 222/325 (68%), Gaps = 3/325 (0%)
```

```
Query: 5   IYDITIVGGGPVGLFAAFYAGLRGVSVKIIESLSELGGQPAILYPEKKIYDIPGYPVITG 64
           +YDITI+GGGP GLFAAFY G+R   VKIIES+ +LGGQ A LYPEK IYD+ G+P +
Sbjct: 7   LYDITIIGGGPTGLFAAFYGGMRQAKVKIIESMPQLGGQLAALYPEKYIYDVAGFPKVKA 66

Query: 65  RELIDKHIEQLERFKDSIEICLKEEVLSFEK-VDDVFTIQTDKDQHLSRAIVFACGNGAF 123
           ++L++    Q E+F  +I    L++ V +  K  DD FTI+TDK+ H S+AI+   G GAF
Sbjct: 67  QDLVNDLKRQAEQFNPTI--ALEQSVQNVTKETDDTFTIKTDKETHYSKAIIITAGAGAF 124

Query: 124 APRLLGLENEENYADNNLFYNVTKLEQFAGKHVVICGGGDSAVDWANELDKIAASVAIVH 183
            PR L +E  + Y   NL Y V  L  +AGK+V+I GGGDSAVDWA  L+ +A +V ++H
Sbjct: 125 QPRRLEVEGAKQYEGKNLQYFVNDLNAYAGKNVLISGGGDSAVDWALMLEPVAKNVTLIH 184

Query: 184 RRDAFRAHEHSVDILKASGVRILTPYVPIGLNGDSQRVSSLVVQKVKGDEVIELPLDNLI 243
           RRD FRAHEHSV++L+ S V ILTP+    L+GD +++  + +Q+VKGD V  L +D +I
Sbjct: 185 RRDKFRAHEHSVELLQKSSVNILTPFAISELSGDGEKIHHVTIQEVKGDAVETLDVDEVI 244

Query: 244 VSFGFSTSNKNLRYWNLDYKRSSINVSSLFETTQEGVYAIGDAANYPGKVELIATGYGEA 303
           V+FGF +S   ++ W L+ +++SI V++  ET   G+YA GD   YPGKV+LIATG+GEA
Sbjct: 245 VNFGFVSSLGPIKGWGLEIEKNSIVVNTKMETNIPGIYAAGDICTYPGKVKLIATGFGEA 304

Query: 304 PVAINQAINYIYPDRDNRVVHSTSL 328
           P A+N A  +I P       HSTSL
Sbjct: 305 PTAVNNAKAFIDPTARVFPGHSTSL 329
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6079> which encodes the amino acid sequence <SEQ ID 6080>. Analysis of this protein sequence reveals the following:

```
      Possible site: 20
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -0.37   Transmembrane    8 -  24 (   8 -  24)

----- Final Results -----
               bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB15201 GB:Z99120 similar to thioredoxin reductase [Bacillus subtilis]
 Identities = 173/328 (52%), Positives = 223/328 (67%), Gaps = 4/328 (1%)

Query: 4   KAYDITIIGGGPIGLFAAFYAGLRGVTVKIIESLSELGGQPAILYPEKMIYDIPAYPSLT 63
           K YDITIIGGGP+GLF AFY G+R  +VKIIESL +LGGQ + LYPEK IYD+  +P +
Sbjct: 6   KVYDITIIGGGPVGLFTAFYGGMRQASVKIIESLPQLGGQLSALYPEKYIYDVAGFPKIR 65

Query: 64  GVELTENLIKQLSRFEDRTTICLKEEVLTFDKVKGG-FSIRTNKAEHFSKAIIIACGNGA 122
             EL  NL +Q+++F+   TICL++ V + +K   G F +   K     K   I  GNGA
Sbjct: 66  AQELINNLKEQMAKFDQ--TICLEQAVESVEKQADGVFKLVQMKKPTTLKRSCITAGNGA 123

Query: 123 FAPRTLGLESEENFADHNLFYNVHQLDQFAGQKVVICGGGDSAVDWALALEDIAESVTVV 182
           F PR L LE+ E +   NL Y V  L +FAG++V I GGGDSAVDWAL LE IA+ V+++
Sbjct: 124 FKPRKLELENAEQYEGKNLHYFVDDLQKFAGRRVAILGGGDSAVDWALMLEPIAKEVSII 183

Query: 183 HRRDAFRAHEHSVELLKASTVNLLTPYVPKALKGIGNLAEKLVIQKVKEDEVLELELDSL 242
           HRRD FRAHEHSVE L AS VN+LTP+VP  L G   + E+LV+++VK D    LE+D L
Sbjct: 184 HRRDKFRAHEHSVENLHASKVNVLTPFVPAELIGEDKI-EQLVLEEVKGDRKEILEIDDL 242

Query: 243 IVSFGFSTSNKNLKNWNLDYKRSSITVSPLFQTSQEGIFAIGDAAAYNGKVDLIATGFGE 302
           IV++GF +S   +KNW LD +++SI V    +T+ EG FA GD   Y GKV+LIA+GFGE
Sbjct: 243 IVNYGFVSSLGPIKNWGLDIEKNSIVVKSTMETNIEGFFAAGDICTYEGKVNLIASGFGE 302

Query: 303 APTAVNQAINYIYPDRDNRVVHSTSLID 330
           APTAVN A  Y+ P    + +HSTSL +
Sbjct: 303 APTAVNNAKAYMDPKARVQPLHSTSLFE 330
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 242/324 (74%), Positives = 279/324 (85%)

Query: 6   YDITIVGGGPVGLFAAFYAGLRGVSVKIIESLSELGGQPAILYPEKKIYDIPGYPVITGR 65
           YDITI+GGGP+GLFAAFYAGLRGV+VKIIESLSELGGQPAILYPEK IYDIP YP +TG
Sbjct: 6   YDITIIGGGPIGLFAAFYAGLRGVTVKIIESLSELGGQPAILYPEKMIYDIPAYPSLTGV 65

Query: 66  ELIDKHIEQLERFKDSIEICLKEEVLSFEKVDDVFTIQTDKDQHLSRAIVFACGNGAFAP 125
           EL +  I+QL RF+D   ICLKEEVL+F+KV   F+I+T+K +H S+AI+ ACGNGAFAP
Sbjct: 66  ELTENLIKQLSRFEDRTTICLKEEVLTFDKVKGGFSIRTNKAEHFSKAIIIACGNGAFAP 125

Query: 126 RLLGLENEENYADNNLFYNVTKLEQFAGKHVVICGGGDSAVDWANELDKIAASVAIVHRR 185
           R LGLE+EEN+AD+NLFYNV +L+QFAG+ VVICGGGDSAVDWA  L+ IA SV +VHRR
Sbjct: 126 RTLGLESEENFADHNLFYNVHQLDQFAGQKVVICGGGDSAVDWALALEDIAESVTVVHRR 185

Query: 186 DAFRAHEHSVDILKASGVRILTPYVPIGLNGDSQRVSSLVVQKVKGDEVIELPLDNLIVS 245
           DAFRAHEHSV++LKAS V +LTPYVP  L G       LV+QKVK DEV+EL LD+LIVS
Sbjct: 186 DAFRAHEHSVELLKASTVNLLTPYVPKALKGIGNLAEKLVIQKVKEDEVLELELDSLIVS 245

Query: 246 FGFSTSNKNLRYWNLDYKRSSINVSSLFETTQEGVYAIGDAANYPGKVELIATGYGEAPV 305
           FGFSTSNKNL+ WNLDYKRSSI VS LF+T+QEG++AIGDAA Y GKV+LIATG+GEAP
Sbjct: 246 FGFSTSNKNLKNWNLDYKRSSITVSPLFQTSQEGIFAIGDAAAYNGKVDLIATGFGEAPT 305

Query: 306 AINQAINYIYPDRDNRVVHSTSLI 329
           A+NQAINYIYPDRDNRVVHSTSLI
Sbjct: 306 AVNQAINYIYPDRDNRVVHSTSLI 329
```

SEQ ID 6078 (GBS178) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 38 (lane 5; MW 37.4kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 41 (lane 8; MW 62.4kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1962

A DNA sequence (GBSx2071) was identified in *S.agalactiae* <SEQ ID 6081> which encodes the amino acid sequence <SEQ ID 6082>. This protein is predicted to be tRNA methyltransferase (trmD). Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1496(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06198 GB:AP001515 tRNA methyltransferase [Bacillus halodurans]
 Identities = 144/246 (58%), Positives = 186/246 (75%), Gaps = 6/246 (2%)

Query: 2   MKIDILTLFPEMFAPLEHS-IVGKAKERGLLEINYHNFRENAE-KSRHVDDEPYGGGQGM 59
           MKID LTLFPEMF  + HS I+ +A+ERG +     NFRE +E K + VDD PYGGG GM
Sbjct: 1   MKIDFLTLFPEMFQGVLHSSILKQAQERGAVSFRVVNFREYSENKHKKVDDYPYGGGAGM 60

Query: 60  LLRAQPIFDTIDKIDAQKA---RVILLDPAGRTFDQDFAEELSKEDELIFICGHYEGYDE 116
           +L  QP+FD ++ +  + +   RVIL+ P G TF Q  AEEL++ + LI +CGHYEGYDE
Sbjct: 61  VLSPQPLFDAVEDLTKKSSSTPRVILMCPQGETFTQRKAEELAQAEHLILLCGHYEGYDE 120

Query: 117 RIKS-LVTDEVSLGDFVLTGGELAAMTMVDATVRLIPEVIGKETSHQDDSFSSGLLEYPQ 175
           RI+S LVTDE+S+GD+VLTGGEL AM + D+  RL+P V+G ETS Q DSFS+GLLEYPQ
Sbjct: 121 RIRSYLVTDELSIGDYVLTGGELGAMVIADSVTRLLPAVLGNETSAQTDSFSTGLLEYPQ 180
```

```
Query: 176 YTRPYDYLGMTVPDVLMSGHHENIRKWRLEQSLRKTLERRPDLLENYAMTDEERLILEKI 235
           YTRP D+ G  VPDVL+SGHH+NI +WR EQSL++TLERRPDLLE   +T+EE+ +L+ I
Sbjct: 181 YTRPADFRGWKVPDVLLSGHHQNIERWRKEQSLKRTLERRPDLLEGRKLTEEEQELLDSI 240

Query: 236 KTEIER 241
           + + E+
Sbjct: 241 RKQQEK 246
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6083> which encodes the amino acid sequence <SEQ ID 6084>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2705(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 195/240 (81%), Positives = 224/240 (93%)

Query: 2   MKIDILTLFPEMFAPLEHSIVGKAKERGLLEINYHNFRENAEKSRHVDDEPYGGGQGMLL 61
            MKIDILTLFPEMFAPLEHSIVGKAKE+GLL+I+YHNFR+ AEK+RHVDDEPYGGGQGMLL
Sbjct: 1   MKIDILTLFPEMFAPLEHSIVGKAKEKGLLDIHYHNFRDYAEKARHVDDEPYGGGQGMLL 60

Query: 62  RAQPIFDTIDKIDAQKARVILLDPAGRTFDQDFAEELSKEDELIFICGHYEGYDERIKSL 121
            RAQPIFDTI++I+A+K R+ILLDPAG+ F Q +AEEL+ E+ELIFICGHYEGYDERIK+L
Sbjct: 61  RAQPIFDTIEQIEAKKPRIILLDPAGKPFTQAYAEELALEEELIFICGHYEGYDERIKTL 120

Query: 122 VTDEVSLGDFVLTGGELAAMTMVDATVRLIPEVIGKETSHQDDSFSSGLLEYPQYTRPYD 181
            VTDE+SLGDFVLTGGELAAMTMVDATVRLIP+V+GKE+SHQDDSFSSGLLEYPQYTRPYD
Sbjct: 121 VTDEISLGDFVLTGGELAAMTMVDATVRLIPQVLGKESSHQDDSFSSGLLEYPQYTRPYD 180

Query: 182 YLGMTVPDVLMSGHHENIRKWRLEQSLRKTLERRPDLLENYAMTDEERLILEKIKTEIER 241
            Y GMTVPDVLMSGHHE IR WRLE+SL+KT  RRPDLLE+Y  ++EER +L+KIK  +++
Sbjct: 181 YRGMTVPDVLMSGHHERIRLWRLEESLKKTYLRRPDLLEHYNFSEEERKLLDKIKEALDQ 240
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1963

A DNA sequence (GBSx2072) was identified in *S.agalactiae* <SEQ ID 6085> which encodes the amino acid sequence <SEQ ID 6086>. This protein is predicted to be 16S rRNA processing protein. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.71    Transmembrane   32 -  48 (  32 -  52)

----- Final Results -----
               bacterial membrane --- Certainty=0.2084(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9381> which encodes amino acid sequence <SEQ ID 9382> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB13475 GB:Z99112 similar to hypothetical proteins [Bacillus subtilis]
```

```
 Identities = 88/174 (50%), Positives = 128/174 (72%), Gaps = 1/174 (0%)

Query: 54  VTMEYFNVGKIVNTQGLQGEMRVLSVTDFVEERFKKGQVLALFDEKNQFVMDIEIASHRK 113
            +T  +FNVGKIVNT G++GE+RV+S TDF EER+K G  L LF +     +++ + +HR
Sbjct: 1   MTKRWFNVGKIVNTHGIKGEVRVISKTDFAEERYKPGNTLYLFMDGRNEPVEVTVNTHRL 60

Query: 114 QKNFDIIKFKGMYHINDIEKYKGFTLKVAEDQLSDLKDGEFYYHEIIGLDVYEGE-ELIG 172
             K F +++FK   ++N++E+ K   +KV E++L +L +GEFY+HEIIG +V+  E ELIG
Sbjct: 61  HKQFHLLQFKERQNLNEVEELKNAIIKVPEEELGELNEGEFYFHEIIGCEVFTEEGELIG 120

Query: 173 KIKEILQPGANDVWVVERHGKRDLLLPYIPPVVLEVDLSNQRVQVELMEGLDDE 226
            K+KEIL PGANDVWV+ R GK+D L+PYI  VV  +D+  +++++ELMEGL DE
Sbjct: 121 KVKEILTPGANDVWVIGRKGKKDALIPYIESVVKHIDVREKKIEIELMEGLIDE 174
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6087> which encodes the amino acid sequence <SEQ ID 6088>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2787(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 133/172 (77%), Positives = 153/172 (88%)

Query: 56  MEYFNVGKIVNTQGLQGEMRVLSVTDFVEERFKKGQVLALFDEKNQFVMDIEIASHRKQK 115
            MEYFNVGKIVNTQGLQGEMRVLSV+DF EERFKKG  LALFD+K++FV ++ I SHRKQK
Sbjct: 1   MEYFNVGKIVNTQGLQGEMRVLSVSDFAEERFKKGSQLALFDDKDRFVQEVTIVSHRKQK 60

Query: 116 NFDIIKFKGMYHINDIEKYKGFTLKVAEDQLSDLKDGEFYYHEIIGLDVYEGEELIGKIK 175
            +FDIIKFK MYHIN IEKYKG+TLKV++D   DL++GEFYYH+IIG+ VYE + LIG +K
Sbjct: 61  HFDIIKFKDMYHINAIEKYKGYTLKVSKDNQGDLQEGEFYYHQIIGMAVYEKDVLIGHVK 120

Query: 176 EILQPGANDVWVVERHGKRDLLLPYIPPVVLEVDLSNQRVQVELMEGLDDED 227
            EILQPGANDVW+V+R GKRDLLLPYIPPVVL VD+ N+RV VELMEGLDDED
Sbjct: 121 EILQPGANDVWIVKRQGKRDLLLPYIPPVVLNVDVPNKRVDVELMEGLDDED 172
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1964

A DNA sequence (GBSx2073) was identified in *S.agalactiae* <SEQ ID 6089> which encodes the amino acid sequence <SEQ ID 6090>. This protein is predicted to be similar to E. coli ykfC (11). Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3488(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9909> which encodes amino acid sequence <SEQ ID 9910> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC38715 GB:AF030367 maturase-related protein [Streptococcus pneumoniae]
 Identities = 366/425 (86%), Positives = 396/425 (93%)

Query: 12  MSELLDKILSRNNMLEAYKQVKSNKGSAGINGVTIEQMDDYLHQNWRETKQLIKERSYKP 71
            MS+LLDKILSR NMLEAY QVKSNKGSAGI+G+TIE+MD+YL QNWR TK+LIK+R YKP
Sbjct: 1   MSKLLDKILSRENMLEAYNQVKSNKGSAGIDGMTIEEMDNYLRQNWRLTKELIKQRKYKP 60

Query: 72  QPVLRVEIPKPNGGVRNLGIPTAMDRMIQQAIVQVLSPLCEKHFSEYSYGFRPNRSCETA 131
            QPVL+VEIPKP+GG+R LGIPT MDRMIQQAIVQV+SP+CE HFS+ SYGFRPNRSCE A
Sbjct: 61  QPVLKVEIPKPDGGIRQLGIPTVMDRMIQQAIVQVMSPICEPHFSDTSYGFRPNRSCEKA 120

Query: 132 IVQLLEYLNDGYEWIVDIDLEKFFDTVPQDRLMSLVHNIIQDGDTESLIRKYLHSGVVIN 191
            I++LLEYLNDGYEWIVDIDLEKFFDTVPQDRLMSLVHNII+DGDTESLIRKYLHSGV+IN
Sbjct: 121 IMKLLEYLNDGYEWIVDIDLEKFFDTVPQDRLMSLVHNIIEDGDTESLIRKYLHSGVIIN 180

Query: 192 GQRHKTLVGTPQGGNLSPLLSNIMLNELDKGLEKRGLRFVRYADDCVITVGSEAAAKRVM 251
            GQR+KTLVGTPQGGNLSPLLSNIMLNELDK LEKRGLRFVRYADDCVITVGSEAAAKRVM
Sbjct: 181 GQRYKTLVGTPQGGNLSPLLSNIMLNELDKELEKRGLRFVRYADDCVITVGSEAAAKRVM 240

Query: 252 HSVSSYIEKRLGLKVNMTKTKIVRPNKLKYLGFGFWKSPKGWKCRPHQDSVQSFKRKLKQ 311
            +SVS +IEKRLGLKVNMTKTKI RP +LKYLGFGFWKS  GWK RPHQDSV+ FK KLK+
Sbjct: 241 YSVSRFIEKRLGLKVNMTKTKITRPRELKYLGFGFWKSSDGWKSRPHQDSVRRFKLKLKK 300

Query: 312 LTMRKWSIDLITRIERLNWVIRGWINYFSLGNMKSIMTQIDERLRTRIRVIIWKQWKKKA 371
            LT RKWSIDL  RIE+LN  IRGWINYFSLGNMKSI+  IDERLRTR+R+IIWKQWKKK+
Sbjct: 301 LTQRKWSIDLTRRIEQLNLSIRGWINYFSLGNMKSIVASIDERLRTRLRMIIWKQWKKKS 360

Query: 372 KRLWGLLKLGVARWIADKVSGWGDHYQLVAQKSVLKRAISKPALAKRGLVSCLDYYLERH 431
            +RLWGLLKLGV +WIADKVSGWGDHYQLVAQKSVLKRAISKP L KRGLVSCLDYYLERH
Sbjct: 361 RRLWGLLKLGVPKWIADKVSGWGDHYQLVAQKSVLKRAISKPVLEKRGLVSCLDYYLERH 420

Query: 432 ALKVS 436
            ALKVS
Sbjct: 421 ALKVS 425
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1965

A DNA sequence (GBSx2074) was identified in *S.agalactiae* <SEQ ID 6091> which encodes the amino acid sequence <SEQ ID 6092>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -0.37   Transmembrane    7 -   23 (    7 -   23)

----- Final Results -----
              bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 821> which encodes the amino acid sequence <SEQ ID 822>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -2.87   Transmembrane 1157 -1173 (1157 -1174)

----- Final Results -----
              bacterial membrane --- Certainty=0.2147(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 1031/1064 (96%), Positives = 1042/1064 (97%)

Query: 1     MRKKQKLPFDKLAIALISTSILLNAQSDIKANTVTEDTPATEQAVEPPQPIAVSEESPSS 60
             +RKKQKLPFDKLAIAL+STSILLNAQSDIKANTVTEDTPATEQAVE PQP AVSEE+PSS
Sbjct: 1     LRKKQKLPFDKLAIALMSTSILLNAQSDIKANTVTEDTPATEQAVETPQPTAVSEEAPSS 60

Query: 61    KETKTSQTPSDVGETVADDANDLAPQAPAKTADTPATSKATIRDLNDPSHVKTLQEKAGK 120
             KETKT QTP D  ET+ADDANDLAPQAPAKTADTPATSKATIRDLNDPS VKTLQEKAGK
Sbjct: 61    KETKTPQTPDDAEETIADDANDLAPQAPAKTADTPATSKATIRDLNDPSQVKTLQEKAGK 120

Query: 121   GVGTVVAVIDAGFDKNHEAWRLTDKTKARYQSKENLEKAKKEHGITYGEWVNDKVAYYHD 180
             G GTVVAVIDAGFDKNHEAWRLTDKTKARYQSKE+LEKAKKEHGITYGEWVNDKVAYYHD
Sbjct: 121   GAGTVVAVIDAGFDKNHEAWRLTDKTKARYQSKEDLEKAKKEHGITYGEWVNDKVAYYHD 180

Query: 181   YSKDGKNAVDQEHGTHVSGILSGNAPSEMKEPYRLEGAMPEAQLLLMRVEIVNGLADYAR 240
             YSKDGK AVDQEHGTHVSGILSGNAPSE KEPYRLEGAMPEAQLLLMRVEIVNGLADYAR
Sbjct: 181   YSKDGKTAVDQEHGTHVSGILSGNAPSETKEPYRLEGAMPEAQLLLMRVEIVNGLADYAR 240

Query: 241   NYAQAIRDAVNLGAKVINMSFGNAALAYANLPDETKKAFDYAKSKGVSIVTSAGNDSSFG 300
             NYAQAI DAVNLGAKVINMSFGNAALAYANLPDETKKAFDYAKSKGVSIVTSAGNDSSFG
Sbjct: 241   NYAQAIIDAVNLGAKVINMSFGNAALAYANLPDETKKAFDYAKSKGVSIVTSAGNDSSFG 300

Query: 301   GKPRLPLADHPDYGVVGTPAAADSTLTVASYSPDKQLTETATVKTDDHQDKEMPVLSTNR 360
             GK RLPLADHPDYGVVGTPAAADSTLTVASYSPDKQLTETATVKT D QDKEMPVLSTNR
Sbjct: 301   GKTRLPLADHPDYGVVGTPAAADSTLTVASYSPDKQLTETATVKTADQQDKEMPVLSTNR 360

Query: 361   FEPNKAYDYAYANRGTKEDDFKDVEGKIALIERGDIDFKDKIANAKKAGAVGVLIYDNQD 420
             FEPNKAYDYAYANRG KEDDFKDV+GKIALIERGDIDFKDKIANAKKAGAVGVLIYDNQD
Sbjct: 361   FEPNKAYDYAYANRGMKEDDFKDVKGKIALIERGDIDFKDKIANAKKAGAVGVLIYDNQD 420

Query: 421   KGFPIELPNVDQMPAAFISRRDGLLLKDNPQKTITFNATPKVLPTASGTKLSRFSSWGLT 480
             KGFPIELPNVDQMPAAFISR+DGLLLK+NPQKTITFNATPKVLPTASGTKLSRFSSWGLT
Sbjct: 421   KGFPIELPNVDQMPAAFISRKDGLLLKENPQKTITFNATPKVLPTASGTKLSRFSSWGLT 480

Query: 481   ADGNIKPDIAAPGQDILSSVANNKYAKLSGTSMSAPLVAGIMGLLQKQYETQYPDMTPSE 540
             ADGNIKPDIAAPGQDILSSVANNKYAKLSGTSMSAPLVAGIMGLLQKQYETQYPDMTPSE
Sbjct: 481   ADGNIKPDIAAPGQDILSSVANNKYAKLSGTSMSAPLVAGIMGLLQKQYETQYPDMTPSE 540

Query: 541   RLDLAKKVLMSSATALYDEDEKAYFSPRQQGAGAVDAKKASAATMYVTDKDNTSSKVHLN 600
             RLDLAKKVLMSSATALYDEDEKAYFSPRQQGAGAVDAKKASAATMYVTDKDNTSSKVHLN
Sbjct: 541   RLDLAKKVLMSSATALYDEDEKAYFSPRQQGAGAVDAKKASAATMYVTDKDNTSSKVHLN 600

Query: 601   NVSDKFEVTVTVHNKSDKPQELYYQVTVQTDKVDGKHFALAPKALYETSWQKITIPANSS 660
             NVSDKFEVTVTVHNKSDKPQELYYQ TVQTDKVDGK FALAPKALYETSWQKITIPANSS
Sbjct: 601   NVSDKFEVTVTVHNKSDKPQELYYQATVQTDKVDGKLFALAPKALYETSWQKITIPANSS 660

Query: 661   KQVTVPIDASRFSKDLLAQMKNGYFLEGFVRFKQDPTKEELMSIPYIGFRGDFGNLSALE 720
             KQVT+PID S+FSKDLLA MKNGYFLEGFVRFKQDPTKEELMSIPYIGFRGDFGNLSALE
Sbjct: 661   KQVTIPIDVSQFSKDLLAPMKNGYFLEGFVRFKQDPTKEELMSIPYIGFRGDFGNLSALE 720

Query: 721   KPIYDSKDGSSYYHEANSDAKDQLDGDGLQFYALKNNFTALTTESNPWTIIKAVKEGVEN 780
             KPIYDSKDGSSYYHEANSDAKDQLDGDGLQFYALKNNFTALTTESNPWTIIKAVKEGVEN
Sbjct: 721   KPIYDSKDGSSYYHEANSDAKDQLDGDGLQFYALKNNFTALTTESNPWTIIKAVKEGVEN 780

Query: 781   IEDIESSEITETIFAGTFAKQDDDSHYYIHRHANGKPYAAISPNGDGNRDYVQFQGTFLR 840
             IEDIESSEITETIFAGTFAKQDDDSHYYIHRHANGKPYAAISPNGDGNRDYVQFQGTFLR
Sbjct: 781   IEDIESSEITETIFAGTFAKQDDDSHYYIHRHANGKPYAAISPNGDGNRDYVQFQGTFLR 840

Query: 841   NAKNLVAEVLDKEGNVVWTSEVTEQVVKNYNNDLASTLGSTRFEKTRWDGKNKDGKVVAN 900
             NAKNLVAEVLDKEGNVVWTSEVTEQVVKNYNNDLASTLGSTRFEKTRWDGK+KDGKVVAN
Sbjct: 841   NAKNLVAEVLDKEGNVVWTSEVTEQVVKNYNNDLASTLGSTRFEKTRWDGKDKDGKVVAN 900

Query: 901   GTYTYRVRYTPISSGAKEQHTDFDVIVDNTTPEVATSATFSTEDSRLTLASKPKTSQPVY 960
             GTYTYRVRYTPISSGAKEQHTDFDVIVDNTTPEVATSATFSTED RLTLASKPKTSQPVY
Sbjct: 901   GTYTYRVRYTPISSGAKEQHTDFDVIVDNTTPEVATSATFSTEDRRLTLASKPKTSQPVY 960
```

```
Query: 961  RERIAYTYMDEDLPTTEYISPNEDGTFTLPEEAETMEGATVPLKMSDFTYVVEDMAGNIT 1020
            RERIAYTYMDEDLPTTEYISPNEDGTFTLPEEAETMEGATVPLKMSDFTYVVEDMAGNIT
Sbjct: 961  RERIAYTYMDEDLPTTEYISPNEDGTFTLPEEAETMEGATVPLKMSDFTYVVEDMAGNIT 1020

Query: 1021 YTPVTKLLEGHSNKPEQDGSDQAPDKKPEAKPEQDGSGQTPDKK 1064
            YTPVTKLLEGHSNKPEQDGSDQAPDKKPE KPEQDGSGQ PDKK
Sbjct: 1021 YTPVTKLLEGHSNKPEQDGSDQAPDKKPETKPEQDGSGQAPDKK 1064
```

A related GBS gene <SEQ ID 8941> and protein <SEQ ID 8942> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 10
McG: Discrim Score:      5.69
GvH: Signal Score (-7.5): -3.33
     Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 1 value:  -0.37 threshold:  0.0
   INTEGRAL    Likelihood = -0.37   Transmembrane    7 -  23 (   7 -  23)
   PERIPHERAL  Likelihood =  2.81     508
 modified ALOM score:   0.57

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

SEQ ID 8942 (GBS276) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 46 (lane 2; MW 123kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 60 (lane 5; MW 46.5kDa).

The GBS276-His fusion product was purified (Figure 206, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 296), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1966

A DNA sequence (GBSx2075) was identified in *S.agalactiae* <SEQ ID 6093> which encodes the amino acid sequence <SEQ ID 6094>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.4286(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1967

A DNA sequence (GBSx2076) was identified in *S.agalactiae* <SEQ ID 6095> which encodes the amino acid sequence <SEQ ID 6096>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.15   Transmembrane   19 -  35 (  11 -  39)

----- Final Results -----
               bacterial membrane --- Certainty=0.5458(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9911> which encodes amino acid sequence <SEQ ID 9912> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 6096 (GBS654) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 142 (lane 8 & 10; MW 51.2kDa + lane 9; MW 27kDa). Purified GBS654-GST is shown in Figure 245, lane 11.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1968

A DNA sequence (GBSx2077) was identified in *S.agalactiae* <SEQ ID 6097> which encodes the amino acid sequence <SEQ ID 6098>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4174(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9913> which encodes amino acid sequence <SEQ ID 9914> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF27324 GB:AF178424 unknown [Lactococcus lactis]
 Identities = 26/75 (34%), Positives = 45/75 (59%), Gaps = 4/75 (5%)

Query: 11 MAFEPKNSELTKVLKES-LDEEKKEIFSSEMNIRDFERTKQYQFTLQPSVRKKIDRLSKE 69
           MAF+  + ++  VL  S L + K E+      I   E  K Y FTL+PSV++ +++L+++
Sbjct: 1  MAFDVDDKKVKTVLSNSSLAKSKVEL---PKKIESEENKKSYSFTLEPSVKEGLEKLAEK 57

Query: 70 KGYRSASSFINDFFK 84
           + Y++ S F+ND  K
Sbjct: 58 QNYKNTSQFLNDLIK 72
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1969

A DNA sequence (GBSx2078) was identified in *S.agalactiae* <SEQ ID 6099> which encodes the amino acid sequence <SEQ ID 6100>. This protein is predicted to be ParA. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF27325 GB:AF178424 ParA [Lactococcus lactis]
 Identities = 49/104 (47%), Positives = 72/104 (69%)

Query: 22  LSERLEEFKTEAFDFKTRASYVTAKLFFLGNMIKHNTNSSKELIRSLKNDKSVLAMIPHK 81
           L ERL+ FK E  D +TR +Y+TA  +F+GN I+HNT SS+E    +  DK  +AMIP K
Sbjct: 157 LIERLQNFKDEVIDARTRETYITAIPYFVGNRIRHNTKSSREFSEKISQDKGTIAMIPEK 216

Query: 82  ELFNRSTLDKKSLSYMMSDKELYSRDSKFFKEIDFTFRKITDKL 125
           ELFNRSTLD   L  M  DK++++ +  F+++++F F +IT+K+
Sbjct: 217 ELFNRSTLDGVPLVEMEKDKDVFNSNKVFYEKLNFAFNEITNKI 260
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1970

A DNA sequence (GBSx2079) was identified in *S.agalactiae* <SEQ ID 6101> which encodes the amino acid sequence <SEQ ID 6102>. This protein is predicted to be transposase (orfA). Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2830(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1971

A DNA sequence (GBSx2080) was identified in *S.agalactiae* <SEQ ID 6103> which encodes the amino acid sequence <SEQ ID 6104>. This protein is predicted to be transposase (orfB). Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2618(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB90834 GB:AJ250837 putative transposase [Streptococcus dysgalactiae]
 Identities = 242/259 (93%), Positives = 249/259 (95%)

Query: 1   MCRWLNMPHSSYYYQAVESVSETEFEETIKRIFLDSESRYGSRKIKICLNNEGITLSRRR 60
            MCRWLN+P SSYYY+AVE VSE E EE+IK IFL+S++RYGSRKIKICLNNEGITLSRRR
Sbjct: 1   MCRWLNIPRSSYYYKAVEPVSEAELEESIKAIFLESKARYGSRKIKICLNNEGITLSRRR 60

Query: 61  IRRIMKRLNLVSVYQKATFKPHSRGKNEAPIPNHLDRQFKQERPLQALVTDLTYVRVGNR 120
            IRRIMKRLNLVSVYQKATFKPHSRGKNEAPIPNHLDRQFK ERPLQALVTDLTYVRVGNR
Sbjct: 61  IRRIMKRLNLVSVYQKATFKPHSRGKNEAPIPNHLDRQFKPERPLQALVTDLTYVRVGNR 120

Query: 121 WAYVCLIIDLYNREIIGLSLGWHKTAELVKQAIQSIPYALTKVKMFHSDRXKEFDNQLID 180
            WAYVCLIIDLYNREIIGLSLGWHKTAELVKQAIQSIPY LTKVKMFHSDR KEF+NQLID
Sbjct: 121 WAYVCLIIDLYNREIIGLSLGWHKTAELVKQAIQSIPYPLTKVKMFHSDRGKEFNNQLID 180

Query: 181 EILEAFGITRSLSQAGCPYDNAVAESTYRAFKIEFVYQETFQLLEELALKTKDYVHWWNY 240
            EILEAFGITRSLSQAGCPYDNAVAESTYRAFKIEFVYQETFQ LEELALKTK YVHWWNY
Sbjct: 181 EILEAFGITRSLSQAGCPYDNAVAESTYRAFKIEFVYQETFQSLEELALKTKAYVHWWNY 240

Query: 241 HRIHGSLNYQTPMTKRLIA 259
            HRIHGSLNYQTPMTKRLIA
Sbjct: 241 HRIHGSLNYQTPMTKRLIA 259
```

There is also homology to SEQ ID 32.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1972

A DNA sequence (GBSx2081) was identified in *S.agalactiae* <SEQ ID 6105> which encodes the amino acid sequence <SEQ ID 6106>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3325(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1973

A DNA sequence (GBSx2082) was identified in *S.agalactiae* <SEQ ID 6107> which encodes the amino acid sequence <SEQ ID 6108>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
               bacterial cytoplasm --- Certainty=0.4442(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9917> which encodes amino acid sequence <SEQ ID 9918> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD44095 GB:AF115103 orf359 gp [Streptococcus thermophilus
           bacteriophage Sfi21]
 Identities = 92/357 (25%), Positives = 162/357 (44%), Gaps = 33/357 (9%)

Query: 45  RKNQYGKTFETMKEAYDELVRIKYEFANKVSLENYNMTFENYMNKIYLRAYKQK-VQSVT 103
            RK +    F T  EA      ++ +  + V+++  ++T  +Y  K +   YK+  V  +T
Sbjct: 24  RKPKTKGGFRTKSEAIKAAAEMELKLQDNVNVDE-DITLYDYF-KQWCEVYKKPTVSKIT 81

Query: 104 YKTALPHHKLFIQYFGLKPLKAITPRDCEAFRLHIIENYSENYAKNLWSRF----KACMG 159
            YK  +   +    +FG K LK+IT  + +     ++ +Y++ +A++   RF    KAC+
Sbjct: 82  YKAYINSQRKIELFFGDKKLKSITATEYQ----RVLNSYAKTHAQDTVERFNVHVKACIE 137

Query: 160 YAERLGYISNMPCKALD---NPRGKHPETPFWTYAEFQTFIKSFDLHDYEELQRFTAIWL 216
             A   GYI    CK        +G+  ET F    E++  I  ++   + E   + A+++
Sbjct: 138 MAVHEGYIKRNFCKFAKINAKNKGRDIETKFLEVEEYERLI--YETSKHPEYASYAALYI 195

Query: 217 YYMTGVRVSEGLSLCWEDIDFDKKFLKVHTTLEKDENGNWYRKDQTKTPAGERLIELDDI 276
               TG+R +E L L  +DI  D   L V+ T +   N  +     TKT +  R I LDD
Sbjct: 196 IAKTGIRFAECLGLTVDDIKRDTGMLSVNKTWDYKNNTGFM---PTKTKSSIREIPLDDE 252

Query: 277 TIEVLQVWRKNQFANQDTDFIISRFGDPFCKSTICRIIKRKAQQVGVPVITGKGLRHSHA 336
             I  +     +Q    D   I+    +     T+ +I+ R+        +    LRH++A
Sbjct: 253 FINFI-----DQLPPTDDGRILPSLSNNAVNKTLRKIVGRE--------VRVHSLRHTYA 299

Query: 337 SYLINVLKKDILYVARRMGHADKSTTLNTYSHWFNALDKTVSEEITQNIKSAGLDSI 393
            SYLI     D++ V++ +GH + + TL  Y+H          E+I Q     G +++
Sbjct: 300 SYLI-AHDIDLISVSQVLGHENLNITLEVYAHQLQEQKSRNDEKIKQMWTECGRNAL 355
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6109> which encodes the amino acid sequence <SEQ ID 6110>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.5549(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 111/127 (87%), Positives = 119/127 (93%)

Query: 242 LKVHTTLEKDENGNWYRKDQTKTPAGERLIELDDITIEVLQVWRKNQFANQDTDFIISRF 301
            LKVHTTLEKDENGNWYRKDQTKTPAGERLIELDD+TI VL+ WR+NQ  N DTDFIISRF
Sbjct: 1   LKVHTTLEKDENGNWYRKDQTKTPAGERLIELDDVTIVVLENWRRNQVVNTDTDFIISRF 60

Query: 302 GDPFCKSTICRIIKRKAQQVGVPVITGKGLRHSHASYLINVLKKDILYVARRMGHADKST 361
            G+PFCKSTICR+IK KAQ +GVPVITGKGLRHS+ASYLINVLKKDILYVA+ MGHADKST
Sbjct: 61  GEPFCKSTICRVIKHKAQSIGVPVITGKGLRHSYASYLINVLKKDILYVAKCMGHADKST 120

Query: 362 TLNTYSH 368
            TLNTYSH
Sbjct: 121 TLNTYSH 127
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1974

A DNA sequence (GBSx2083) was identified in *S.agalactiae* <SEQ ID 6111> which encodes the amino acid sequence <SEQ ID 6112>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3299(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1975

A DNA sequence (GBSx2084) was identified in *S.agalactiae* <SEQ ID 6113> which encodes the amino acid sequence <SEQ ID 6114>. This protein is predicted to be repressor protein-related protein. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2721(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9919> which encodes amino acid sequence <SEQ ID 9920> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC98432 GB:L29324 repressor protein [Streptococcus pneumoniae]
 Identities = 38/65 (58%), Positives = 52/65 (79%), Gaps = 1/65 (1%)

Query: 2  MYRRLRDLREDNDFTQKYVAEK-LSFTHSAYSKIERGERILSADVIIKLSNLYNVSTDYL 60
           M +R+RDLRED+D TQ+YVA+  L+ T SAYSK+E G R++S D +IKL++ YNVS DYL
Sbjct: 1  MLKRIRDLREDDDLTQEYVAKTILNCTRSAYSKMESGTRLISIDDLIKLADFYNVSLDYL 60

Query: 61 LGQTD 65
           +G+ D
Sbjct: 61 VGRVD 65
```

There is also homology to SEQ ID 582.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1976

A DNA sequence (GBSx2085) was identified in *S.agalactiae* <SEQ ID 6115> which encodes the amino acid sequence <SEQ ID 6116>. This protein is predicted to be relaxase. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3160(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC98434 GB:L29324 relaxase [Streptococcus pneumoniae]
 Identities = 223/417 (53%), Positives = 310/417 (73%), Gaps = 5/417 (1%)

Query: 1   MVITKHYAVHGKKYRRQLIKYILDPKKTRNLSLISDFGMSNYLDFPDYVELVKMYQNNFL 60
           MVITKH+A+HGK YR +LIKYIL+P KT+NL+L+SDFGM NYLDFP Y ELVKMY +NFL
Sbjct: 1   MVITKHFAIHGKNYRSKLIKYILNPSKTKNLTLVSDFGMRNYLDFPSYKELVKMYNDNFL 60

Query: 61  SNDQLYDSRFDRQEKKQQKIHAHHIIQSFSPEDKLSPEEINRIGYETIKELIGGQYKFIV 120
           SND LY+ R DRQE  Q+KIH+HHIIQSFSP+D L+PE+INRIGYE  KEL GG+++FIV
Sbjct: 61  SNDTLYEFRHDRQEVNQRKIHSHHIIQSFSPDDHLTPEQINRIGYEAAKELTGGRFRFIV 120

Query: 121 ATHVDQDHCHNHIIINSINSQSQKKLKWDYALERNLQMISDRISKVAGAKIIPPKRYSHR 180
           ATHVD+ H HNHII+NSI+  S KK  WDY  E NL+M+SDR+SK+AGAKII   RYSHR
Sbjct: 121 ATHVDKGHIHNHIILNSIDQNSDKKFLWDYKAEHNLRMVSDRLSKIAGAKII-ENRYSHR 179

Query: 181 DYEVYRRSNHKYELKQRLFFLMEHSIDFNDFMQKAEQLNVKIDFSRKHSRFFMTDRNMKQ 240
            YEVYR++N+KYE+KQR++FL+E+S +F D  +KA+ L++KIDF  KH  +FMTD NMKQ
Sbjct: 180 QYEVYRKTNYKYEIKQRVYFLIENSKNFEDLKKKAKALHLKIDFRHKHVTYFMTDSNMKQ 239

Query: 241 VIQGDKLNKREPYSKEYFQRYFAKKKIELILEFLLLRSNSFDDLVEKARLLGLELKSKKK 300
           V++  KL++++PY++ YF++ F +++I  ILEFLL +  + ++L+++A + GL++  K+K
Sbjct: 240 VVRDSKLSRKQPYNETYFEKKFVQREIINILEFLLPKMKNMNELIQRAEVFGLKIIPKEK 299

Query: 301 TIDFVLSDGKSCISIPNKSLRKKNLYDTTYFDSYFKEHDVFEVLHNNEVKIEFEKFETQQ 360
            + F   DG   I +  + L K NLY  +YF  YF   +   VL N  +   + + +  +
Sbjct: 300 HVLFEF-DG---IKLAEQELVKSNLYSVSYFQDYFNNKNETFVLDNKNLVELYNEEKIIK 355

Query: 361 LSEILTVEEITEAYETYKTKRDAVHEFEVEITEEQIEKIVLDGLFVKVWMGIGQEGL 417
             E+ + E + ++Y+ +K  RDAVHEFEVE+   QIE++V  G+++KV  GI ++ L
Sbjct: 356 EKELPSEEMVWKSYQDFKRNRDAVHEFEVELNLNQIEEVVEHGIYIKVQFGIDKKDL 412
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6117> which encodes the amino acid sequence <SEQ ID 6118>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3114(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 103/218 (47%), Positives = 170/218 (77%)

Query: 393 EEQIEKIVLDGLFVKVWMGIGQEGLIFIPNHQLNILEQENKKQYQVFIRETSSYFIYHKE 452
           E QIE+++ + +++KV   + Q GLIFIPN+QL+I ++EN K+Y+V+IRET+ +FIY+KE
Sbjct: 2   EHQIERLIAEDIYIKVSFSVKQSGLIFIPNYQLDIRKEENHKKYKVYIRETAQFFIYNKE 61
```

```
Query: 453 DSEMNRFMKGRDLIRQLTFDNKSLPYKRRISLVSLQQKIEEINLLMTLNIQNKSFLELKD 512
            SE+NR+M+G +LI QLT D+KS+P +RR ++ +L++KIEEI+LL+ L+ +NK + ++KD
Sbjct: 62  ASELNRYMRGHELICQLTNDSKSIPKRRRQTIDTLKKKIEEISLLIELDTENKPYQDIKD 121

Query: 513 ELVGDIAQLDIELTNLQDKNTTLNKMAEVVVNLQSDNQDTKQLAKYECSKMNLSQNVTIG 572
           ++V D+AQLD+ +T LQD    LNK+AEV++NL +++ + ++LA+Y+ +KMNL+  + I
Sbjct: 122 DIVKDMAQLDLTITELQDHIAHLNKVAEVLLNLNNNDIENRRLARYDYAKMNLTAAIKIE 181

Query: 573 QIESEIEMIQNQLDNKIEEYENAVRKLDEYVRVLNMDK 610
           ++E EIE  QN+L+  I+EYE  VR+L+++  +L+  K
Sbjct: 182 EVEKEIETSQNELNISIDEYEYLVRRLEKFGEILSDSK 219
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1977

A DNA sequence (GBSx2086) was identified in *S.agalactiae* <SEQ ID 6119> which encodes the amino acid sequence <SEQ ID 6120>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4006(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC98436 GB:L29324 unknown [Streptococcus pneumoniae]
 Identities = 53/115 (46%), Positives = 77/115 (66%), Gaps = 2/115 (1%)

Query: 5   VREIRKEVNFSIEEYQQIQNFMEQEGYEQFSPFARGKLLKIDHQPSQQLEEWIKYLQHQK 64
           +R IRK+   +  E +QI + M ++G + FS F R  LL  D Q  +Q+E+W    + QK
Sbjct: 5   IRSIRKQFRLTETEEKQILDLMREKGDDNFSDFLRKSLLLSDGQ--KQMEKWFNLWKKQK 62

Query: 65  VEQIYRDVHEILVLAKLSQSVTMEHLEIILTCIKDLMKEIEVTIPLSYSFKDKYM 119
           +EQI RDVHE+ ++AK +  VT EH+ I+LTCI++L+KE+E T PLS  F +KYM
Sbjct: 63  LEQISRDVHEVFIIAKTNHQVTHEHVSILLTCIQELIKEVEKTGPLSEDFCNKYM 117
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1978

A DNA sequence (GBSx2087) was identified in *S.agalactiae* <SEQ ID 6121> which encodes the amino acid sequence <SEQ ID 6122>. This protein is predicted to be TnpA. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2935(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC82523 GB:AF027768 TnpA [Serratia marcescens]
```

```
 Identities = 176/413 (42%), Positives = 243/413 (58%), Gaps = 18/413 (4%)

Query: 26  MMFKVEAVGPPERCPECGFD-KLYKHSSRNQLIMDLPIRLKRVGLHLNRRRYKCRECGST 84
            M F+V+ V  P  C ECG   +  +   R+    DLPI  KRV L + RRRY CR C +T
Sbjct: 1   MHFQVD-VPDPIACEECGVQGEFVRFGKRDVPYRDLPIHGKRVTLWVVRRRYTCRACKTT 59

Query: 85  IS------VDEKRSMTKRLLKSIQEQSMSKTFVEVAESVGVDEKTIRNVFKDYVALKERE 138
                    VD  R MT RL + ++++S +  +  VA   G+DEKT+R++F        R
Sbjct: 60  FRPQLPEMVDGFR-MTLRLHEYVEKESFNHPYTFVAAQTGLDEKTVRDIFNARAEFLGRW 118

Query: 139 YQFETPKWLGIDEIHIIRRPRLVLTNIERRTIYDIKPNRNKETVIQRLSEISDRTYIEYV 198
            ++FETP+ LGIDE+++ +R R +LTNIE RT+ D+   R ++ V   L ++ DR  +E V
Sbjct: 119 HRFETPRILGIDELYLNKRYRCILTNIEERTLLDLLATRRQDVVTNYLMKLKDRQKVEIV 178

Query: 199 TMDMWKPYKDAVNTILPQAKVVVDKFHVVRMANQALDNVRKSLKAHMSQKERRTLMRERF 258
            +MDMW PY+ AV  +LPQA++VVDKFHVVRMAN AL+ VRK L+  +   + RTL  +R
Sbjct: 179 SMDMWNPYRAAVKAVLPQARIVVDKFHVVRMANDALERVRKGLRKELKPSQSRTLKGDRK 238

Query: 259 ILLKRKHDLNERESFLLDTWLGNLPALKEAYELKEEFYWIWDTPDPDEGHLRYSQWRHRC 318
            ILLKR H++++RE  +++TW G  P L  AYE KE FY IWD     +      +W
Sbjct: 239 ILLKRAHEVSDRERLIMETWTGAFPQLLAAYEHKERFYGIWDATTRLQAEAALDEW-IAT 297

Query: 319 MSSNSKDAYKDLVRAVDNWHVEIFNYF--DKRLTNAYTESINSIIRQVERMGRGYSFDAL 376
            +    K+ + DLVRAV NW  E   YF  D  +TNAYTESIN + +   R GRGYSF+ +
Sbjct: 298 IPKGQKEVWSDLVRAVGNWREETMTYFETDMPVTNAYTESINRLAKDKNREGRGYSFEVM 357

Query: 377 RAKILFNEKLHKKRKPRFNSSAFNKAMLYDTFNWYEVNDHDITDNLGVDFSTL 429
            RA++L+  K HKK+ P    S F K  +      Y + D     N GVD ST+
Sbjct: 358 RARMLYTTK-HKKKAPTAKVSPFYKKTI-----GYGLPDFAEELNYGVDLSTI 404
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1979

A DNA sequence (GBSx2088) was identified in *S.agalactiae* <SEQ ID 6123> which encodes the amino acid sequence <SEQ ID 6124>. This protein is predicted to be mercuric reductase. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2115(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA70224 GB:Y09024 mercuric reductase [Bacillus cereus]
 Identities = 412/546 (75%), Positives = 484/546 (88%)

Query: 1   MNKFKVNISGMTCTGCEKHVESALEKIGAKNIESSYRRGEAVFELPDDIEVESAIKAIDE 60
            M K++V++ GMTCTGCE+HV  ALE +GA  IE  +RRGEAVFELP+ + VE+A KAI +
Sbjct: 1   MKKYRVDVQGMTCTGCEEHVAVALENMGATGIEVDFRRGEAVFELPNALGVETAKKAISD 60

Query: 61  ANYQAGEIEEVSSLENVALINEDNYDLLIIGSGAAAFSSAIKAIEYGAKVGMIERGTVGG 120
            A YQ G+ EEV S E V L NE +YD +IIGSG AAFSSAI+A++YGAKV MIERGT+GG
Sbjct: 61  AKYQPGKAEEVQSQEMVQLGNEGDYDYIIIGSGGAAFSSAIEAVKYGAKVAMIERGTIGG 120

Query: 121 TCVNIGCVPSKTLLRAGEINHLSKDNPFIGLQTSAGEVDLASLITQKDKLVSELRNQKYM 180
            TCVNIGCVPSKTLLRAGEINHL+K+NPF+GL TSAGEVDLA LI QK++LV+ELRN KY+
Sbjct: 121 TCVNIGCVPSKTLLRAGEINHLAKNNPFVGLHTSAGEVDLAPLIKQKNELVTELRNSKYV 180
```

```
Query: 181 DLIDEYNFDLIKGEAKFVDASTVEVNGTKLSAKRFLIATGASPSLPQISGLEKMDYLTST 240
           DLID+Y F+LI+GEAKFVD  TVEVNG  +SAKRFLIATGASP+ P I GL ++DYLTST
Sbjct: 181 DLIDDYGFELIEGEAKFVDEKTVEVNGAPISAKRFLIATGASPAKPNIPGLNEVDYLTST 240

Query: 241 TLLELKKIPKRLTVIGSGYIGMELGQLFHHLGSEITLMQRSERLLKEYDPEISESVEKAL 300
           +LLELKK+PKRL VIGSGYIGMELGQLFH+LGSE+TL+QRSERLLKEYDPEISESVEK+L
Sbjct: 241 SLLELKKVPKRLVVIGSGYIGMELGQLFHNLGSEVTLIQRSERLLKEYDPEISESVEKSL 300

Query: 301 IEQGINLVKGATFERVEQSGEIKRVYVTVNGSREVIESDQLLVATGRKPNTDSLNLSAAG 360
           +EQGINLVKGAT+ER+EQ+G+IK+V+V VNG + +IE+DQLLVATGR PNT +LNL AAG
Sbjct: 301 VEQGINLVKGATYERIEQNGDIKKVHVEVNGKKRIIEADQLLVATGRTPNTATLNLRAAG 360

Query: 361 VETGKNNEILINDFGQTSNEKIYAAGDVTLGPQFVYVAAYEGGIITDNAIGGLNKKIDLS 420
           VE G   EI+I+D+ +T+N +IYAAGDVTLGPQFVYVAAY+GG+   NAIGGLNKK++L
Sbjct: 361 VEIGSRGEIIIDDYSRTTNTRIYAAGDVTLGPQFVYVAAYQGGVAAPNAIGGLNKKLNLE 420

Query: 421 VVPAVTFTNPTVATVGLTEEQAKEKGYDVKTSVLPLDAVPRAIVNRETTGVFKLVADAET 480
           VVP VTFT P +ATVGLTE+QAKE GY+VKTSVLPLDAVPRA+VNRETTGVFKLVAD++T
Sbjct: 421 VVPGVTFTAPAIATVGLTEQQAKENGYEVKTSVLPLDAVPRALVNRETTGVFKLVADSKT 480

Query: 481 LKVLGVHIVSENAGDVIYAASLAVKFGLTIEDLTETLAPYLTMAEGLKLVALTFDKDISK 540
           +KVLG H+V+ENAGDVIYAA+LAVKFGLT++D+ ETLAPYLTMAEGLKL ALTFDKDISK
Sbjct: 481 MKVLGAHVVAENAGDVIYAATLAVKFGLTVDDIRETLAPYLTMAEGLKLAALTFDKDISK 540

Query: 541 LSCCAG 546
           LSCCAG
Sbjct: 541 LSCCAG 546
```

There is also homology to SEQ ID 1820.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1980

A DNA sequence (GBSx2089) was identified in *S.agalactiae* <SEQ ID 6125> which encodes the amino acid sequence <SEQ ID 6126>. This protein is predicted to be regulatory protein. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4529(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA83973 GB:AF138877 mercury resistance operon negative
           regulator MerR1 [Bacillus sp. RC607]
 Identities = 83/129 (64%), Positives = 104/129 (80%)

Query: 1   MIYRISEFADKCGVNKETIRYYERKNLLQEPHRTEAGYRIYSYDDVKRVGFIKRIQEFGF 60
           M +RI E ADKCGVNKETIRYYER  L+ EP RTE GYR+YS   V R+ FIKR+QE GF
Sbjct: 1   MKFRIGELADKCGVNKETIRYYERLGLIPEPERTEKGYRMYSQQTVDRLHFIKRMQELGF 60

Query: 61  SLSEIYKLLGVVDKDEVRCQDMFEFVSKKQKEVQKQIEDLKRIETMLDDLKQRCPDEKKL 120
           +L+EI KLLGVVD+DE +C+DM++F   K +++Q++IEDLKRIE ML DLK+RCP+ K +
Sbjct: 61  TLNEIDKLLGVVDRDEAKCRDMYDFTILKIEDIQRKIEDLKRIERMLMDLKERCPENKDI 120

Query: 121 HSCPIIETL 129
           + CPIIETL
Sbjct: 121 YECPIIETL 129
```

There is also homology to SEQ ID 1712.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1981

A DNA sequence (GBSx2090) was identified in *S.agalactiae* <SEQ ID 6127> which encodes the amino acid sequence <SEQ ID 6128>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.86   Transmembrane   80 -  96 (  78 - 100)

----- Final Results -----
               bacterial membrane --- Certainty=0.4142(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8943> and protein <SEQ ID 8944> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:    -13.52
GvH: Signal Score (-7.5): -6.14
     Possible site: 44
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value:  -7.86 threshold:  0.0
   INTEGRAL    Likelihood = -7.86   Transmembrane   80 -  96 (  78 - 100)
   PERIPHERAL  Likelihood =  1.80     136
 modified ALOM score:   2.07

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4142(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF02021(439 - 666 of 1080)
GP|451734|gb|AAA18975.1||U05143(9  -  46  of  46)  envelope  glycoprotein  {Simian
immunodeficiency  virus}  GP|451744|gb|AAA18980.1||U05148  envelope  glycoprotein  {Simian
immunodeficiency virus}
%Match = 3.2
%Identity = 38.5  %Similarity = 64.1
Matches = 15  Mismatches = 13  Conservative Sub.s = 10

336       366       396       426       456       486       516       546
RIPVQFKGCDDYYNENVGYPLSRINLEHYLTEGGVLYFVVYSKDVSPTVTYASLTPKVIKNVLPASDKKKRIKKKEDIFL
                                             :|| |  : ||:|::||:  |:
                                    WGLTGNAGTTPTATTTTTTPRVVENVINESN-------------
                                             10        20        30
```

```
576       606       636       666        696       726       756        786
LFWMAIIAKLLILPYPALQTSYKSRPCLRRSSLRKLTQIPFSIVTKVGNTNMKSITAFLQVKAYILPCLAKGPARIMV*W
                         ||:: :|   | | |
-------------------------PCIKDNSCAGLEQEP
                                 40
```

SEQ ID 8944 (GBS415) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 79 (lane 3; MW 21.2kDa).

Example 1982

A DNA sequence (GBSx2092) was identified in *S.agalactiae* <SEQ ID 6129> which encodes the amino acid sequence <SEQ ID 6130>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3402(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes.*

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1983

A DNA sequence (GBSx2093) was identified in *S.agalactiae* <SEQ ID 6131> which encodes the amino acid sequence <SEQ ID 6132>. This protein is predicted to be ATPase. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.08   Transmembrane  324 - 340 ( 317 - 343)
   INTEGRAL    Likelihood = -5.73   Transmembrane  662 - 678 ( 660 - 690)
   INTEGRAL    Likelihood = -5.41   Transmembrane  350 - 366 ( 346 - 378)
   INTEGRAL    Likelihood = -3.40   Transmembrane   94 - 110 (  93 - 110)
   INTEGRAL    Likelihood = -2.87   Transmembrane  681 - 697 ( 680 - 699)
   INTEGRAL    Likelihood = -1.38   Transmembrane  148 - 164 ( 148 - 164)

----- Final Results -----
               bacterial membrane --- Certainty=0.5034(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA22858 GB:M90750 cadmium-efflux ATPase [Bacillus firmus]
 Identities = 486/725 (67%), Positives = 584/725 (80%), Gaps = 18/725 (2%)

Query: 1   MSRGKAKQSEKEMKAYRVQGFTCTNCAAIFENNVKELPGVQDAKVNFGASKVYVKGTTTI 60
           MS  KA  SE+EMKAYRVQGFTC NCA  FE NVK+L GV+DAKVNFGASK+ V G  TI
Sbjct: 1   MSDQKAITSEQEMKAYRVQGFTCANCAGKFEKNVKQLSGVEDAKVNFGASKIAVYGNATI 60

Query: 61  EELEKAGAFENLKIRDEKEQRVGGE-----------PFWKQKENIKVYISALLLVVSWFL 109
           EELEKAGAFENLK+  EK  R   +           PF+K K +  +Y S LL+   +
Sbjct: 61  EELEKAGAFENLKVTPEKSARQASQEVKEDTKEDKVPFYK-KHSTLLYAS-LLITFGYLS 118

Query: 110 GEQYGEEHVLPTIGYAASILIGGYSLFIKGLKNLRRLNFDMNTLMTIAIIGAAIIGEWGE 169
               GEE+++ T+ + AS+ IGG SLF  GL+NL R  FDM TLMT+A+IG AIIGEW E
```

```
Sbjct: 119 SYVNGEENIVTTLLFLASMFIGGLSLFKVGLQNLLRFEFDMKTLMTVAVIGGAIIGEWAE 178

Query: 170 GATVVILFAISEALERYSMDKARQSIESLMDIAPKEALIRRGNEEMMIHVDEIQVGDIMI 229
            A VVILFAISEALER+SMD+ARQSI SLMDIAPKEAL++R  +E+MIHVD+I VGDIMI
Sbjct: 179 VAIVVILFAISEALERFSMDRARQSIRSLMDIAPKEALVKRNGQEIMIHVDDIAVGDIMI 238

Query: 230 VKPGQKLAMDGIVVKGTSTLNQAAITGESVPVTKITNDEVFAGTLNEEGLLEVKVTKRVE 289
           VKPGQK+AMDG+VV G S +NQ AITGESVPV K  ++EVFAGTLNEEGLLEV++TK VE
Sbjct: 239 VKPGQKIAMDGVVVSGYSAVNQTAITGESVPVEKTVDNEVFAGTLNEEGLLEVEITKLVE 298

Query: 290 DTTLSKIIHLVEEAQAERAPSQAFVDKFAKYYTPAIVILALLIAVVPPL-FGGDWSQWIY 348
           DTT+SKIIHLVEEAQ ERAPSQAFVDKFAKYYTP I+I+A L+A+VPPL F G W  WIY
Sbjct: 299 DTTISKIIHLVEEAQGERAPSQAFVDKFAKYYTPIIMIIATLVAIVPPLFFDGSWETWIY 358

Query: 349 QGLAVLVVGCPCALVVSTPVAVVTAIGNAAKNGVLIKGGIHLEAAGHLKAIAFDKTGTLT 408
           QGLAVLVVGCPCALV+STP+++V+AIGNAAK GVL+KGG++LE  G LKAIAFDKTGTLT
Sbjct: 359 QGLAVLVVGCPCALVISTPISIVSAIGNAAKKGVLVKGGVYLEEMGALKAIAFDKTGTLT 418

Query: 409 KGIPAVTD--IVTYGRNENELITITSAIEKGSQHPLASAIMRKAEENGLKFNEVTVEDFQ 466
           KG+PAVTD  ++    NE EL++I +A+E  SQHPLASAIM+KAEE  + +++V VEDF
Sbjct: 419 KGVPAVTDYNVLNKQINEKELLSIITALEYRSQHPLASAIMKKAEEENITYSDVQVEDFS 478

Query: 467 SITGKGVKAKINNEMYYVGSQNLFEE-LHGSISSDKKEKIADMQTQGKTVMVLGTEKEIL 525
           SITGKG+K  +N   YY+GS  LF+E L      D ++ +  +Q QGKT M++GTEKEIL
Sbjct: 479 SITGKGIKGIVNGTTYYIGSPKLFKELLTNDFDKDLEQNVTTLQNQGKTAMIIGTEKEIL 538

Query: 526 SFIAVADEMRESSKEVIGKLNNMGI-ETVMLTGDNQRTATAIGKQVGVSDIKADLLPEDK 584
           + IAVADE+RESSKE++ KL+ +GI +T+MLTGDN+ TA AIG QVGVSDI+A+L+P+DK
Sbjct: 539 AVIAVADEVRESSKEILQKLHQLGIKKTIMLTGDNKGTANAIGGQVGVSDIEAELMPQDK 598

Query: 585 LNFIKELREKHQSVGMVGDGVNDAPALAASTVGVAMGGAGTDTALETADIALMSDDLSKL 644
           L+FIK+LR ++ +V MVGDGVNDAPALAASTVG+AMGGAGTDTALETAD+ALM DDL KL
Sbjct: 599 LDFIKQLRSEYGNVAMVGDGVNDAPALAASTVGIAMGGAGTDTALETADVALMGDDLRKL 658

Query: 645 PYTIKLSRKALAIIKQNITFSLAIKLVALLLVMPGWLTLWIAIFADMGATLLVTLNSLRL 704
           P T+KLSRK L IIK NITF++AIK +A LLV+PGWLTLWIAI +DMGATLLV LN LRL
Sbjct: 659 PSTVKLSRKTLNIIKANITFAIAIKFIASLLVIPGWLTLWIAILSDMGATLLVALNGLRL 718

Query: 705 LKIKE 709
           +++KE
Sbjct: 719 MRVKE 723
```

There is also homology to SEQ ID 3506.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1984

A DNA sequence (GBSx2094) was identified in *S.agalactiae* <SEQ ID 6133> which encodes the amino acid sequence <SEQ ID 6134>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0779(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1985

A DNA sequence (GBSx2095) was identified in *S.agalactiae* <SEQ ID 6135> which encodes the amino acid sequence <SEQ ID 6136>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.92   Transmembrane  123 - 139 ( 115 - 145)
   INTEGRAL    Likelihood = -6.74   Transmembrane  172 - 188 ( 167 - 190)
   INTEGRAL    Likelihood = -1.81   Transmembrane   80 -  96 (  80 -  96)

----- Final Results -----
              bacterial membrane --- Certainty=0.4567(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9923> which encodes amino acid sequence <SEQ ID 9924> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 4216.

A related GBS gene <SEQ ID 8945> and protein <SEQ ID 8946> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:     -6.41
GvH: Signal Score (-7.5): -2.23
     Possible site: 58
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 3 value:  -8.92 threshold:  0.0
   INTEGRAL    Likelihood = -8.92   Transmembrane  123 - 139 ( 115 - 145)
   INTEGRAL    Likelihood = -6.74   Transmembrane  172 - 188 ( 167 - 190)
   INTEGRAL    Likelihood = -1.81   Transmembrane   80 -  96 (  80 -  96)
   PERIPHERAL  Likelihood =  2.92      46
 modified ALOM score:   2.28

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.4567(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1986

A DNA sequence (GBSx2096) was identified in *S.agalactiae* <SEQ ID 6137> which encodes the amino acid sequence <SEQ ID 6138>. This protein is predicted to be histidine rich P type ATPase (HRA-1) (copB). Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-13.37   Transmembrane  318 - 334 ( 307 - 345)
   INTEGRAL    Likelihood = -5.84   Transmembrane  347 - 363 ( 335 - 364)
   INTEGRAL    Likelihood = -5.15   Transmembrane   88 - 104 (  86 - 112)
```

```
INTEGRAL    Likelihood = -5.04   Transmembrane  651 - 667 ( 649 - 669)
INTEGRAL    Likelihood = -4.30   Transmembrane  156 - 172 ( 155 - 173)
INTEGRAL    Likelihood = -4.30   Transmembrane  669 - 685 ( 668 - 690)
INTEGRAL    Likelihood = -3.03   Transmembrane   62 -  78 (  60 -  80)

----- Final Results -----
            bacterial membrane --- Certainty=0.6349(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA62113 GB:U16658 histidine rich P type ATPase [Escherichia
           coli]
 Identities = 598/731 (81%), Positives = 651/731 (88%), Gaps = 36/731 (4%)

Query: 1   MRNNKKHSSHSHHNHGDIDHSKHDHNEMEHSQMDHS------------------------ 36
           MRNNK+HSSHSHHNHGD++HSKHDHNEMEHSQMDHS
Sbjct: 1   MRNNKQHSSHSHHNHGDMEHSKHDHNEMEHSQMDHSAMGHCAMGGHAHHHHGDMDHSKHD 60

Query: 37  ------------NMDHSEMDHGAMGGHAHHHHGSFKEIFLKSLPLGIAILLITPMMDIQL 84
                        MD+SEMDHGAMGGHAHHHHGSFK+IFLKSLPLGIAILLITP+M IQL
Sbjct: 61  HNEMKHSQMDHSKMDYSEMDHGAMGGHAHHHHGSFKDIFLKSLPLGIAILLITPLMGIQL 120

Query: 85  PFQIIFPYADVVAAVLATILYIFGGKPFYMGAKDEFNSKAPGMMSLITLGITVSYAYSVY 144
           PFQIIFPYADVVAAVLATILYIFGGKPF MGAKDEFNSK PGMMSLITLGITVSYAYSVY
Sbjct: 121 PFQIIFPYADVVAAVLATILYIFGGKPFLMGAKDEFNSKVPGMMSLITLGITVSYAYSVY 180

Query: 145 AVAARYVTGEHVMDFFFEFTTLILIMLLGHWIEMKALGEAGDAQKALAELVPKDAHVVLE 204
           AVAARYVTGE VMDFFFEFTTLILIMLLGHWIEMKALGEAG+AQKALAELVPKDAHVVLE
Sbjct: 181 AVAARYVTGEPVMDFFFEFTTLILIMLLGHWIEMKALGEAGNAQKALAELVPKDAHVVLE 240

Query: 205 DDSIETRPVSELQIGDVIRVQAGENVPADGIIIRGESRVNEALVTGESKPIEKKTGDEVI 264
           DDSIETRPV++LQ+GD+IRVQAGENVPADG I RGESRVNEALVTGESKPIEK  GDEVI
Sbjct: 241 DDSIETRPVADLQVGDLIRVQAGENVPADGTIQRGESRVNEALVTGESKPIEKNPGDEVI 300

Query: 265 GGSTNGGGVLYVEIKQTGDQSFISQVQTLISQAQSQPSRAENVAQKVASWLFYIAVVVAL 324
           GGSTNG GVLYVEIKQTGD+SFISQVQTLISQAQSQPSRAEN+AQKVA WLFYIAV+ AL
Sbjct: 301 GGSTNGDGVLYVEIKQTGDKSFISQVQTLISQAQSQPSRAENLAQKVAGWLFYIAVIAAL 360

Query: 325 IALLIWTIIADLPTAVIFTVTALVIACPHALGLAIPLVVSRSTSLGASRGLLVKNREALE 384
           IAL+IW +IAD+PTAVIFTVT LVIACPHALGLAIPLV +RSTSLGASRGLLVK+R+ALE
Sbjct: 361 IALVIWMVIADVPTAVIFTVTTLVIACPHALGLAIPLVTARSTSLGASRGLLVKDRDALE 420

Query: 385 LTTKADVMVLDKTGTLTTGEFKVLDVTVLSDKYSEEEITGLLAGIEAGSSHPIAQSIVNH 444
           LTT ADVMVLDKTGTLTTGEFKVLDV + +DKY+++EI  LL+GIE GSSHPIAQSI+++
Sbjct: 421 LTTNADVMVLDKTGTLTTGEFKVLDVELFNDKYTKDEIVALLSGIEGGSSHPIAQSIISY 480

Query: 445 AEAKGIKSVSFDSIEIVSGAGIEGEANGHHYQLISQKAYGKALRMDIPKGATLSILVENN 504
           AE +GI+ VSFDSI+++SGAG+EG+ANGH YQLISQKAYG+ L MDIPKGAT+S+LVEN+
Sbjct: 481 AEQQGIRPVSFDSIDVMSGAGVEGQANGHRYQLISQKAYGRNLDMDIPKGATISVLVEND 540

Query: 505 EAIGAVALGDELKETSRNLIEVLKKYGIEPLMATGDNEEAAQGVAEVLGIQYQANQSPED 564
           EAIGAVALGDELK TS++LI+ LKK  I+P+MATGDNE+AAQG AE+LGI Y ANQSP+D
Sbjct: 541 EAIGAVALGDELKPTSKDLIQALKKNKIQPIMATGDNEKAAQGAAEILGIDYLANQSPQD 600

Query: 565 KYKLVESMKNQNKTVIMVGDGVNDAPSLALADVGIAIGAGTQVALDSADIILTQSDPGDI 624
           KY+LVE +K + K VIMVGDGVNDAPSLALADVGIAIGAGTQVALDSADIILTQ  PGDI
Sbjct: 601 KYELVEKLKAEGKKVIMVGDGVNDAPSLALADVGIAIGAGTQVALDSADIILTQYSPGDI 660

Query: 625 ESFIELANKTTRKMKQNLVWGAGYNFIAIPIAAGLLAPIGITLGPAFGAVLMSLSTVIVA 684
            SFIELA KTTRKMK+NLVWGAGYNFIAIPIAAG+LAPIGITL PA  AVLMSLSTVIVA
Sbjct: 661 ASFIELAQKTTRKMKENLVWGAGYNFIAIPIAAGILAPIGITLSPAVAAVLMSLSTVIVA 720

Query: 685 INAMTLKLEPK 695
           INAMTLKLEPK
Sbjct: 721 INAMTLKLEPK 731
```

There is also homology to SEQ ID 3506.

A related GBS gene <SEQ ID 8947> and protein <SEQ ID 8948> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 7
McG: Discrim Score:    -19.12
GvH: Signal Score (-7.5): -3.71
     Possible site: 27
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 7 value: -13.37 threshold:  0.0
   INTEGRAL    Likelihood =-13.37   Transmembrane  291 - 307 ( 280 - 318)
   INTEGRAL    Likelihood = -5.84   Transmembrane  320 - 336 ( 308 - 337)
   INTEGRAL    Likelihood = -5.15   Transmembrane   61 -  77 (  59 -  85)
   INTEGRAL    Likelihood = -5.04   Transmembrane  624 - 640 ( 622 - 642)
   INTEGRAL    Likelihood = -4.30   Transmembrane  129 - 145 ( 128 - 146)
   INTEGRAL    Likelihood = -4.30   Transmembrane  642 - 658 ( 641 - 663)
   INTEGRAL    Likelihood = -3.03   Transmembrane   35 -  51 (  33 -  53)
   PERIPHERAL  Likelihood =  0.74     103
 modified ALOM score:   3.17

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6349(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF02015(220 - 2304 of 2604)
EGAD|37454|38974(1 - 731 of 731) histidine rich P type ATPase (HRA-1) {Escherichia coli}
GP|643613|gb|AAA62113.1||U16658  histidine  rich  P  type  ATPase  {Escherichia  coli}
PIR|JC2464|JC2464  probable  copper-transporting  ATPase  (EC  3.6.1.-)  HRA-1  -
Enterobacteriaceae spp.
%Match = 67.4
%Identity = 85.9  %Similarity = 93.7
Matches = 598  Mismatches = 43  Conservative Sub.s = 54

162       192       222       252
PFRENYM*C*MRKF*NFKISL*YNKEELKMRNNKKHSSHSHHNHGDI---------------------------------
                             |||||:|||||||||||:
                             MRNNKQHSSHSHHNHGDMEHSKHDHNEMEHSQMDHSAMGHCAMGGHAHHHH
                                     10        20        30        40        50

294       324       354       384       414       444       474       504
---DHSKHDHNEMEHSQMDHSNMDHSEMDHGAMGGHAHHHHGSFKEIFLKSLPLGIAILLITPMMDIQLPFQIIFPYADV
   ||||||||||:||||||| ||:||||||||||||||||||||:|||||||||||||||||:| ||||||||||||||
GDMDHSKHDHNEMKHSQMDHSKMDYSEMDHGAMGGHAHHHHGSFKDIFLKSLPLGIAILLITPLMGIQLPFQIIFPYADV
                  70        80        90       100       110       120       130

534       564       594       624       654       684       714       744
VAAVLATILYIFGGKPFYMGAKDEFNSKAPGMMSLITLGITVSYAYSVYAVAARYVTGEHVMDFFFEFTTLILIMLLGHW
||||||||||||||||| |||||||||| |||||||||||||||||||||||||||||| ||||||||||||||||||||
VAAVLATILYIFGGKPFLMGAKDEFNSKVPGMMSLITLGITVSYAYSVYAVAARYVTGEPVMDFFFEFTTLILIMLLGHW
                 150       160       170       180       190       200       210

774       804       834       864       894       924       954       984
IEMKALGEAGDAQKALAELVPKDAHVVLEDDSIETRPVSELQIGDVIRVQAGENVPADGIIIRGESRVNEALVTGESKPI
||||||||||:|||||||||||||||||||||||||||::||:||:|||||||||||||| | |||||||||||||||||
IEMKALGEAGNAQKALAELVPKDAHVVLEDDSIETRPVADLQVGDLIRVQAGENVPADGTIQRGESRVNEALVTGESKPI
                 230       240       250       260       270       280       290

1014      1044      1074      1104      1134      1164      1194      1224
EKKTGDEVIGGSTNGGGVLYVEIKQTGDQSFISQVQTLISQAQSQPSRAENVAQKVASWLFYIAVVVALIALLIWTIIAD
||  ||||||||||| ||||||||||||:||||||||||||||||||||||:||||| |||||||: |||||:|| :|||
EKNPGDEVIGGSTNGDGVLYVEIKQTGDKSFISQVQTLISQAQSQPSRAENLAQKVAGWLFYIAVIAALIALVIWMVIAD
```

```
                310       320       330       340       350       360       370

1254      1284      1314      1344      1374      1404      1434      1464
LPTAVIFTVTALVIACPHALGLAIPLVVSRSTSLGASRGLLVKNREALELTTKADVMVLDKTGTLTTGEFKVLDVTVLSD
:||||||||| |||||||||||||||| :||||||||||||||:|:|||||| |||||||||||||||||||||| :::|
VPTAVIFTVTTLVIACPHALGLAIPLVTARSTSLGASRGLLVKDRDALELTTNADVMVLDKTGTLTTGEFKVLDVELFND
                390       400       410       420       430       440       450

1494      1524      1554      1584      1614      1644      1674      1704
KYSEEEITGLLAGIEAGSSHPIAQSIVNHAEAKGIKSVSFDSIEIVSGAGIEGEANGHHYQLISQKAYGKALRMDIPKGA
||:::||  ||:||| ||||||||||:::|| :||: ||||||:::||||:||:|||| ||||||||||: | |||||||
KYTKDEIVALLSGIEGGSSHPIAQSIISYAEQQGIRPVSFDSIDVMSGAGVEGQANGHRYQLISQKAYGRNLDMDIPKGA
                470       480       490       500       510       520       530

1734      1764      1794      1824      1854      1884      1914      1944
TLSILVENNEAIGAVALGDELKETSRNLIEVLKKYGIEPLMATGDNEEAAQGVAEVLGIQYQANQSPEDKYKLVESMKNQ
|:|:||||:|||||||||||||| ||::||: |||  |:|:|||||||:|||| ||:||| | |||||:|||:||| :| :
TISVLVENDEAIGAVALGDELKPTSKDLIQALKKNKIQPIMATGDNEKAAQGAAEILGIDYLANQSPQDKYELVEKLKAE
                550       560       570       580       590       600       610

1974      2004      2034      2064      2094      2124      2154      2184
NKTVIMVGDGVNDAPSLALADVGIAIGAGTQVALDSADIILTQSDPGDIESFIELANKTTRKMKQNLVWGAGYNFIAIPI
 | ||||||||||||||||||||||||||||||||||||||||  |||| |||||| |||||||:|||||||||||||||
GKKVIMVGDGVNDAPSLALADVGIAIGAGTQVALDSADIILTQYSPGDIASFIELAQKTTRKMKENLVWGAGYNFIAIPI
                630       640       650       660       670       680       690

2214      2244      2274      2304      2334      2364      2394      2424
AAGLLAPIGITLGPAFGAVLMSLSTVIVAINAMTLKLEPK*NEAGTKKHWLV*PPSRIGSDQLVCCIRKIIDR*IFDKNR
|||:|||||||| ||  |||||||||||||||||||||||
AAGILAPIGITLSPAVAAVLMSLSTVIVAINAMTLKLEPK
                710       720       730
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1987

A DNA sequence (GBSx2097) was identified in *S.agalactiae* <SEQ ID 6139> which encodes the amino acid sequence <SEQ ID 6140>. This protein is predicted to be CopA. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2197(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA40599 GB:X57326 ORF-1 [Thiobacillus ferrooxidans]
 Identities = 26/65 (40%), Positives = 40/65 (61%), Gaps = 2/65 (3%)

Query: 1  MKQEILL--DGVKCAGCANTVQERFSAIEGVESVEVDLATKKAVLESQTEIDTETLNAAL 58
          M Q+I L   G+ CA CA++V++    I G++S +V LAT +A +  Q+ I TE L AA+
Sbjct: 1  MSQKIFLRITGMTCAHCAHSVEKALLGIHGIDSAQVSLATNQAEVFLQSSIPTEALLAAV 60

Query: 59 AETNY 63
           +  Y
Sbjct: 61 TQAGY 65
```

There is also homology to SEQ ID 3510.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1988

A DNA sequence (GBSx2098) was identified in *S.agalactiae* <SEQ ID 6141> which encodes the amino acid sequence <SEQ ID 6142>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3220(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1989

A DNA sequence (GBSx2099) was identified in *S.agalactiae* <SEQ ID 6143> which encodes the amino acid sequence <SEQ ID 6144>. This protein is predicted to be heavy-metal transporting P-type ATPase (b0484). Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.09   Transmembrane  131 - 147 ( 130 - 150)

----- Final Results -----
               bacterial membrane --- Certainty=0.2635(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB01764 GB:U42410 heavy-metal transporting P-type ATPase
           [Proteus mirabilis]
 Identities = 98/153 (64%), Positives = 123/153 (80%)

Query: 2   KAVKALRRRGVEVIMITGDNKRTAKAIAKQVGIDSVLSEVLPEDKAEEVKKLQEAGKKVA 61
            +A+KAL   G++V MITGDNK TAKAIAKQ+GID +++EVLP+ K   +K+L + G KVA
Sbjct: 649 EAIKALHALGLKVAMITGDNKATAKAIAKQLGIDEIVAEVLPDGKVAALKQLSQKGDKVA 708

Query: 62  MVGDGINDAPALAQANVGIAVGSGTDVAIESADIVLMRNDLTAVLTTIDLSHATLRNIKQ 121
             VGDGINDAPALAQA+VG+A+G+GTDVAIE+AD+VLM  DL  V+  I LS AT+RNIKQ
Sbjct: 709 FVGDGINDAPALAQADVGLAIGTGTDVAIEAADVVLMSGDLRGVVDAIALSQATIRNIKQ 768

Query: 122 NLFWAFAYNLVGIPVAMGLLYIFGGLLMSPMLA 154
            NLFW FAYN + IPVA G+LY   G+L+SP+ A
Sbjct: 769 NLFWTFAYNALLIPVAAGMLYPINGMLLSPIFA 801
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3505> which encodes the amino acid sequence <SEQ ID 3506>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.83   Transmembrane  328 - 344 ( 314 - 348)
```

```
INTEGRAL    Likelihood = -7.01   Transmembrane  354 - 370 ( 347 - 377)
INTEGRAL    Likelihood = -3.24   Transmembrane  101 - 117 ( 100 - 117)
INTEGRAL    Likelihood = -2.97   Transmembrane  165 - 181 ( 165 - 185)
INTEGRAL    Likelihood = -2.34   Transmembrane  665 - 681 ( 662 - 684)
INTEGRAL    Likelihood = -2.18   Transmembrane   67 -  83 (  66 -  83)
INTEGRAL    Likelihood = -0.64   Transmembrane  491 - 507 ( 490 - 508)
INTEGRAL    Likelihood = -0.59   Transmembrane  691 - 707 ( 691 - 707)
INTEGRAL    Likelihood = -0.43   Transmembrane  140 - 156 ( 139 - 156)

----- Final Results -----
              bacterial membrane --- Certainty=0.5331(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 92/152 (60%), Positives = 123/152 (80%)

Query: 4   VKALRRRGVEVIMITGDNKRTAKAIAKQVGIDSVLSEVLPEDKAEEVKKLQEAGKKVAMV 63
            V+AL + G+  IM+TGD+  TAKAIA QVGI  V+S+VLP+ KA  +  L+  G+KVAMV
Sbjct: 544 VEALHQLGIHTIMLTGDHDATAKAIASQVGITDVISQVLPDQKAGVIADLRSQGRKVAMV 603

Query: 64  GDGINDAPALAQANVGIAVGSGTDVAIESADIVLMRNDLTAVLTTIDLSHATLRNIKQNL 123
            GDGINDAPALA A++GIA+GSGTD+AIESAD++LM+ D+  ++  + LS  T+R +K+NL
Sbjct: 604 GDGINDAPALAVADIGIAMGSGTDIAIESADVILMKPDMLDLVKAMSLSRVTMRIVKENL 663

Query: 124 FWAFAYNLVGIPVAMGLLYIFGGLLMSPMLAG 155
            FWAF YN++ IPVAMGLL++FGG L++PMLAG
Sbjct: 664 FWAFIYNVLMIPVAMGLLHLFGGPLLNPMLAG 695
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1990

A DNA sequence (GBSx2100) was identified in *S.agalactiae* <SEQ ID 6145> which encodes the amino acid sequence <SEQ ID 6146>. This protein is predicted to be CopY. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2067(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG10085 GB:AF296446 CopY [Streptococcus mutans]
 Identities = 63/139 (45%), Positives = 96/139 (68%)

Query: 8   TSITDAEWEVMRVVWANDLVTSKTVISVLKEKMDWTESTIKTILGRLVEKGVLNTEQEGR 67
            TSI++AEWEVMRVVWA  + +S  +I++L     W+ STIKT++ RL EKG L ++++GR
Sbjct: 2   TSISNAEWEVMRVVWAKQMTSSSEIIAILSRTYCWSASTIKTLITRLSEKGYLTSQRQGR 61

Query: 68  KFIYTANIVEKEAVRDFAEDIFNRICKKKVGNVIGSIIEDHVLSFDDIDRLEKILEIKKS 127
            K+IY++ I E+EA+     ++F+RIC  K   +I  ++E+  ++  DI++LE +L  KK+
Sbjct: 62  KYIYSSLISEEEALEQQVSEVFSRICVTKHQALIRHLVEETPMTLSDIEKLEALLLSKKA 121

Query: 128 FAVEEVDCQCTEGQCDCHE 146
             AV EV C C  GQC C+E
Sbjct: 122 NAVPEVKCNCIVGQCSCYE 140
```

There is also homology to SEQ ID 3502.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1991

A DNA sequence (GBSx2101) was identified in *S.agalactiae* <SEQ ID 6147> which encodes the amino acid sequence <SEQ ID 6148>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2829(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1992

A DNA sequence (GBSx2102) was identified in *S.agalactiae* <SEQ ID 6149> which encodes the amino acid sequence <SEQ ID 6150>. This protein is predicted to be DS RF protein. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-13.21   Transmembrane  142 - 158 ( 136 - 169)
   INTEGRAL    Likelihood = -3.45   Transmembrane   70 -  86 (  66 -  88)
   INTEGRAL    Likelihood = -3.13   Transmembrane  178 - 194 ( 176 - 195)

----- Final Results -----
               bacterial membrane --- Certainty=0.6286(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA26611 GB:L10909 putative [Staphylococcus aureus]
 Identities = 98/204 (48%), Positives = 148/204 (72%), Gaps = 3/204 (1%)

Query: 4   TIISAIGVYISTSIDYLIVLIILFAQLSQNKQKWHIYAGQYLGTGLLVGASLVAAY-VVN 62
           TI++A  VY++T IDYL++LI+LF+Q+ + + K HI+ GQY+GT +++GASL+ A  VVN
Sbjct: 18  TILTATAVYVATGIDYLVILILLFSQVKKGQVK-HIWIGQYIGTAIVIGASLLVAQGVVN 76

Query: 63  FVPEAWMVGLLGLIPIYLGIRFAIVGEGEEEEEEEIIERLEQSKANQLFWTVTLLTIASG 122
            +P+ W++GLLGL+P+YLG++  I GE E+E+E  I+      K NQLF T+  + +AS
Sbjct: 77  LIPQQWVIGLLGLLPLYLGVKIWIKGE-EDEDESSILSLFSSGKFNQLFLTMIFIVLASS 135

Query: 123 GDNLGIYIPYFASLDWSQTLVVLLVFAIGIIIFCELSWVLSSIPLISETIEKYQRIIVPL 182
            D+  IYIPYF +L  S+  +V +VF I + + C +S+ L+S   ISETIEKY+R IVP+
Sbjct: 136 ADDFSIYIPYFTTLSMSEIFIVTIVFLIMVGVLCYVSYRLASFDFISETIEKYERWIVPI 195

Query: 183 VFIPLGLYIMYESGTIETFLNFIL 206
           VFI LG+YI++E+GT    ++F+L
Sbjct: 196 VFIGLGIYILFENGTSNALISFLL 219
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6151> which encodes the amino acid sequence <SEQ ID 6152>. Analysis of this protein sequence reveals the following:

```
     Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-13.16   Transmembrane  143 - 159 ( 135 - 165)
   INTEGRAL    Likelihood = -9.13   Transmembrane   49 -  65 (  43 -  71)
   INTEGRAL    Likelihood = -7.17   Transmembrane   73 -  89 (  72 -  94)
   INTEGRAL    Likelihood = -6.00   Transmembrane   13 -  29 (   9 -  33)
   INTEGRAL    Likelihood = -2.71   Transmembrane  180 - 196 ( 179 - 197)
   INTEGRAL    Likelihood = -0.59   Transmembrane  112 - 128 ( 109 - 128)

----- Final Results -----
               bacterial membrane --- Certainty=0.6265(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF42284 GB:AE002544 cadmium resistance protein [Neisseria
           meningitidis MC58]
 Identities = 201/208 (96%), Positives = 205/208 (97%)

Query: 1   MRCFMIQNVVTSIILYSGTAVDLLIILMLFFAKRKSRKDIINIYLGQFLGSVSLILLSLL 60
            MRCFMIQNVVTSIILYSGTAVDLLIILMLFFAKRKSRKDIINIYLGQFLGSVSLILLSLL
Sbjct: 1   MRCFMIQNVVTSIILYSGTAVDLLIILMLFFAKRKSRKDIINIYLGQFLGSVSLILLSLL 60

Query: 61  FAFVLDYIPSKEILGLLGLIPIFLGLKVLLLGDSDGEAIAKEGLSKDNKNLIFLVAMITF 120
            FAFVLDYIPSKEILGLLGLIPI LG+KVLLLGDSDGEAIAKEGL KDNKNLIFLVAMITF
Sbjct: 61  FAFVLDYIPSKEILGLLGLIPILLGIKVLLLGDSDGEAIAKEGLRKDNKNLIFLVAMITF 120

Query: 121 ASCGADNIGVFVPYFTTLNLANLIVALLTFLVMIYLLVFSAQKLAQVPSVGETLEKYSRW 180
            ASCGADNIGVFVPYFTTLNLANLIVALLTFLVMIYLLVFSAQKLAQVPSVGETLEKYSRW
Sbjct: 121 ASCGADNIGVFVPYFTTLNLANLIVALLTFLVMIYLLVFSAQKLAQVPSVGETLEKYSRW 180

Query: 181 FIAVVYLGLGMYILIENNSFDMLWAVLG 208
            F+AVVYLGLG+YIL+ENNSFDMLW VLG
Sbjct: 181 FVAVVYLGLGIYILVENNSFDMLWTVLG 208
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 71/200 (35%), Positives = 130/200 (64%), Gaps = 4/200 (2%)

Query: 1   MGQTIISAIGVYISTSIDYLIVLIILFAQLSQNKQKWHIYAGQYLGTGLLVGASLVAAYV 60
            M Q ++++I +Y  T++D LI+L++ FA+    K   +IY GQ+LG+  L+  SL+ A+V
Sbjct: 5   MIQNVVTSIILYSGTAVDLLIILMLFFAKRKSRKDIINIYLGQFLGSVSLILLSLLFAFV 64

Query: 61  VNFVPEAWMVGLLGLIPIYLGIRFAIVGEGEEEEEEEIIERLEQSKANQLFWTVTLLTIA 120
            ++++P   ++GLLGLIPI+LG++  ++G+ + E   +  E L +   N +F  V ++T A
Sbjct: 65  LDYIPSKEILGLLGLIPIFLGLKVLLLGDSDGEAIAK--EGLSKDNKNLIF-LVAMITFA 121

Query: 121 S-GGDNLGIYIPYFASLDWSQTLVVLLVFAIGIIIFCELSWVLSSIPLISETIEKYQRII 179
            S G DN+G+++PYF +L+ +  +V LL F + I +    +  L+ +P + ET+EKY R
Sbjct: 122 SCGADNIGVFVPYFTTLNLANLIVALLTFLVMIYLLVFSAQKLAQVPSVGETLEKYSRWF 181

Query: 180 VPLVFIPLGLYIMYESGTIE 199
            + +V++ LG+YI+ E+ + +
Sbjct: 182 IAVVYLGLGMYILIENNSFD 201
```

SEQ ID 6150 (GBS174) was expressed in and purified from *E.coli*. The purified protein is shown in lane 7 of Figure 223.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1993

A DNA sequence (GBSx2103) was identified in *S.agalactiae* <SEQ ID 6153> which encodes the amino acid sequence <SEQ ID 6154>. This protein is predicted to be Pgm. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4324(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB96418 GB:AJ243290 phosphoglucomutase [Streptococcus thermophilus]
 Identities = 65/76 (85%), Positives = 71/76 (92%)

Query: 1  MTYTENLQKWLDFEQLPDYLRQELLSMDEKTKEDAFYTNLEFGTAGMRGYIGAGTNRINI 60
           M+YTEN QKWLDF +LP YLR EL+SMDEKTKEDAFYTNLEFGTAGMRG IGAGTNRINI
Sbjct: 1  MSYTENYQKWLDFAELPAYLRDELVSMDEKTKEDAFYTNLEFGTAGMRGLIGAGTNRINI 60

Query: 61 YVVRQATEGLAKLIET 76
           YVVRQATEGLA+LI++
Sbjct: 61 YVVRQATEGLAQLIDS 76
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6155> which encodes the amino acid sequence <SEQ ID 6156>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4324(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 75/76 (98%), Positives = 75/76 (98%)

Query: 1  MTYTENLQKWLDFEQLPDYLRQELLSMDEKTKEDAFYTNLEFGTAGMRGYIGAGTNRINI 60
           MTYTEN QKWLDFEQLPDYLRQELLSMDEKTKEDAFYTNLEFGTAGMRGYIGAGTNRINI
Sbjct: 1  MTYTENFQKWLDFEQLPDYLRQELLSMDEKTKEDAFYTNLEFGTAGMRGYIGAGTNRINI 60

Query: 61 YVVRQATEGLAKLIET 76
           YVVRQATEGLAKLIET
Sbjct: 61 YVVRQATEGLAKLIET 76
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1994

A DNA sequence (GBSx2104) was identified in *S.agalactiae* <SEQ ID 6157> which encodes the amino acid sequence <SEQ ID 6158>. This protein is predicted to be a membrane protein. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -6.21   Transmembrane   94 - 110 (  93 - 115)
   INTEGRAL    Likelihood = -4.14   Transmembrane  172 - 188 ( 166 - 188)
```

```
   INTEGRAL    Likelihood = -1.97   Transmembrane  130 - 146 ( 129 - 149)
   INTEGRAL    Likelihood = -0.16   Transmembrane   62 -  78 (  62 -  79)

----- Final Results -----
               bacterial membrane --- Certainty=0.3484(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA80247 GB:Z22520 membrane protein [Bacillus acidopullulyticus]
 Identities = 47/185 (25%), Positives = 80/185 (42%), Gaps = 23/185 (12%)

Query: 1   MKKKNKSSNIAIIAIFFAIMLVIHFLSSFIFSFWLVPIKPTLMHIPVIIASIAYGPRIGA 60
           MKK     +I I  +  A+ +++                 T+MHIP II  I  GP +G
Sbjct: 1   MKKSLTVRDIVIAGVLGAVAILLGVTRLGYIPVPTAAGNATIMHIPAIIGGIMQGPVVGL 60

Query: 61  TLGALMGGISVANSSIVLLPTSYLFSPFVENGNFYSLIIALVPRILIGIIPYFVYKLLHN 120
            +GA+ G  S  N+++ L               F   +++++PR+ IG++ + VY  +
Sbjct: 61  IVGAIFGISSFLNATVPL---------------FKDPLVSILPRLFIGVVAWLVYIGIRR 105

Query: 121 R---FGLAISGAIGSLTNTVFVLSGIFIFFSSTYNGNIKLMLAGIISSNSLAEMVIAAII 177
           +     + +S  IG+LTNT  VL+     F       +   +A    +N L E V+  I+
Sbjct: 106 KSEYVAVGLSAFIGTLTNTALVLA--MAVFRHYLTAGVAWTVA---ITNGLPEAVVGTIV 160

Query: 178 VYLTV 182
               V
Sbjct: 161 TLAVV 165
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6159> which encodes the amino acid sequence <SEQ ID 6160>. Analysis of this protein sequence reveals the following:

```
     Possible site: 31
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.97   Transmembrane   18 -  34 (  10 -  41)
   INTEGRAL    Likelihood = -7.43   Transmembrane  170 - 186 ( 160 - 191)
   INTEGRAL    Likelihood = -5.63   Transmembrane   96 - 112 (  94 - 117)
   INTEGRAL    Likelihood = -4.67   Transmembrane  140 - 156 ( 131 - 158)
   INTEGRAL    Likelihood = -3.66   Transmembrane   64 -  80 (  63 -  84)
   INTEGRAL    Likelihood = -0.22   Transmembrane   39 -  55 (  39 -  55)

----- Final Results -----
               bacterial membrane --- Certainty=0.4588(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA80247 GB:Z22520 membrane protein [Bacillus acidopullulyticus]
 Identities = 47/193 (24%), Positives = 86/193 (44%), Gaps = 28/193 (14%)

Query: 8   RKSADISRIAIFFAIMLVIHFVSSLVFNIWPIPI---KPTLVHIPVIIASVLYGPRIGAI 64
           +KS  +  I I   +  V   +        P+P      T++HIP II  ++ GP +G I
Sbjct: 2   KKSLTVRDIVIAGVLGAVAILLGVTRLGYIPVPTAAGNATIMHIPAIIGGIMQGPVVGLI 61

Query: 65  LGGLMGIISVITNTIILLPTNYLFSPFVDHGTFASLIIAIIPRILIGITPYYCYKLIPNQ 124
           +G + GI S +  T+ L               F   +++I+PR+ IG+  +  Y  I  +
Sbjct: 62  VGAIFGISSFLNATVPL---------------FKDPLVSILPRLFIGVVAWLVYIGIRRK 106

Query: 125 FGLIVSGI---IGSLTNTIFVLS-GIFIFFATVFDGNIKALLTAIISSNAIVEMIISAII 180
              +  G+   IG+LTNT  VL+  +F  + T        +   +  +N + E ++  I+
Sbjct: 107 SEYVAVGLSAFIGTLTNTALVLAMAVFRHYLTA------GVAWTVAITNGLPEAVVGTIV 160

Query: 181 TFVLIPTLSRLKR 193
           T  ++    ++ R
Sbjct: 161 TLAVVLAWKQIGR 173
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 121/184 (65%), Positives = 157/184 (84%)

Query: 6   KSSNIAIIAIFFAIMLVIHFLSSFIFSFWLVPIKPTLMHIPVIIASIAYGPRIGATLGAL 65
           KS++I+ IAIFFAIMLVIHF+SS +F+ W +PIKPTL+HIPVIIAS+ YGPRIGA LG L
Sbjct: 9   KSADISRIAIFFAIMLVIHFVSSLVFNIWPIPIKPTLVHIPVIIASVLYGPRIGAILGGL 68

Query: 66  MGGISVANSSIVLLPTSYLFSPFVENGNFYSLIIALVPRILIGIIPYFVYKLLHNRFGLA 125
           MG ISV  ++I+LLPT+YLFSPFV++G F SLIIA++PRILIGI PY+ YKL+ N+FGL
Sbjct: 69  MGIISVITNTIILLPTNYLFSPFVDHGTFASLIIAIIPRILIGITPYYCYKLIPNQFGLI 128

Query: 126 ISGAIGSLTNTVFVLSGIFIFFSSTYNGNIKLMLAGIISSNSLAEMVIAAIIVYLTVPRI 185
           +SG IGSLTNT+FVLSGIFIFF++ ++GNIK +L  IISSN++ EM+I+AII ++ +P +
Sbjct: 129 VSGIIGSLTNTIFVLSGIFIFFATVFDGNIKALLTAIISSNAIVEMIISAIITFVLIPTL 188

Query: 186 LNIK 189
             +K
Sbjct: 189 SRLK 192
```

A related GBS gene <SEQ ID 8949> and protein <SEQ ID 8950> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 5
McG: Discrim Score:     13.42
GvH: Signal Score (-7.5): -1.93
     Possible site: 53
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 2 value:  -6.21 threshold:  0.0
   INTEGRAL    Likelihood = -6.21   Transmembrane   94 - 110 (  93 - 115)
   INTEGRAL    Likelihood = -0.16   Transmembrane   62 -  78 (  62 -  79)
   PERIPHERAL  Likelihood =  1.70     123
 modified ALOM score:   1.74

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.3484(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01561(301 - 723 of 1017)
EGAD|38021|39600(1 - 129 of 183) hypothetical membrane protein {Bacillus acidopullulyticus}
GP|806536|emb|CAA80247.1||Z22520 membrane protein {Bacillus acidopullulyticus}
%Match = 7.6
%Identity = 29.7  %Similarity = 53.9
Matches = 38  Mismatches = 57  Conservative Sub.s = 31

162       192       222       252       282       312       342       372
KKIGYQEIEPRISLLACGDTGQGALADISTILKCIQEVAN*AVNLYTISSLI*GVIMKKKNKSSNIAIIAIFFAIMLVIH
                                                        |||     :| |  :: |: :::
                                                        MKKSLTVRDIVIAGVLGAVAILLG
                                                                10        20

402       432       462       492       522       552       582       612
FLSSFIFSFWLVPIKPTLMHIPVIIASIAYGPRIGATLGALMGGISVANSSIVLLPTSYLFSPFVENGNFYSLIIALVPR
                |:|||| ||  |  || :|  :||: |  |  |::: |               |   :::::||
VTRLGYIPVPTAAGNATIMHIPAIIGGIMQGPVVGLIVGAIFGISSFLNATVPL---------------FKDPLVSILPR
              40        50        60        70                       80

642       663       693       723       753       783       813       843
ILIGIIPYFVY---KLLHNRFGLAISGAIGSLTNTVFVXSGIFIFFSSTYNGNIKLMLAGIISXNSLAEMVIAAIIVYLT
::||:: ::||   :       : :|  ||:|||| :| :
LFIGVVAWLVYIGIRRKSEYVAVGLSAFIGTLTNTALVLAMAVFRHYLTAGVAWTVAITNGLPEAVVGTIVTLAVVLAWK
        100       110       120       130       140       150       160
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1995

A DNA sequence (GBSx2105) was identified in *S.agalactiae* <SEQ ID 6161> which encodes the amino acid sequence <SEQ ID 6162>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence (or aa 1-18)

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0165(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC44502 GB:U48885 DNA/pantothenate metabolism flavoprotein
           [Streptococcus mutans]
 Identities = 101/145 (69%), Positives = 122/145 (83%)

Query: 1   MIKRITLAVTGSISAYKAADLTSQLTKIGYDVHIIMTQAATEFITPLTLQVLSKNPIHLD 60
           M K+I LAV+GSI+AYKAADL+ QLTK+GY V++ MT AA +FI PLTLQVLSKNP++ +
Sbjct: 1   MTKKILLAVSGSIAAYKAADLSHQLTKLGYHVNVFMTNAAKQFIPPLTLQVLSKNPVYSN 60

Query: 61  VMDEHNPKIINHIELAKRTDLFIVAPASANTIAHLAYGFADNIVTSVALAMPDETPKLIA 120
           VM E +P++INHI LAK+ DLF++ PASANT+AHLA+GFADNIVTSVALA+P E PK  A
Sbjct: 61  VMKEDDPQVINHIALAKQADLFLLPPASANTLAHLAHGFADNIVTSVALALPLEVPKFFA 120

Query: 121 PAMNTKMYHNTITQRNIDILKKIGY 145
           PAMNTKMY N ITQ NI +LKK GY
Sbjct: 121 PAMNTKMYENPITQSNITLLKKFGY 145
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6163> which encodes the amino acid sequence <SEQ ID 6164>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0076(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 146/178 (82%), Positives = 155/178 (87%)

Query: 1   MIKRITLAVTGSISAYKAADLTSQLTKIGYDVHIIMTQAATEFITPLTLQVLSKNPIHLD 60
           M K ITLAV+GSISAYKAADLTSQLTKIGYDVHIIMTQAAT+FITPLTLQVLSKN IHLD
Sbjct: 1   MTKHITLAVSGSISAYKAADLTSQLTKIGYDVHIIMTQAATQFITPLTLQVLSKNAIHLD 60

Query: 61  VMDEHNPKIINHIELAKRTDLFIVAPASANTIAHLAYGFADNIVTSVALAMPDETPKLIA 120
           VMDEH+PK+INHIELAKRTDLFIVAPASANTIAHLAYGFADN+VTSVALA+P  TPKLIA
Sbjct: 61  VMDEHDPKVINHIELAKRTDLFIVAPASANTIAHLAYGFADNLVTSVALALPATTPKLIA 120

Query: 121 PAMNTKMYHNTITQRNIDILKKIGYQEIEPRISLLACGDTGQGALADISTILKCIQEV 178
           PAMNTKMY N ITQ NI  L  IG+ EI P+ SLLACGD G GALADI  IL  I +
Sbjct: 121 PAMNTKMYQNPITQENIKRLSTIGFTEIPPKSSLLACGDKGPGALADIDVILATIDTI 178
```

SEQ ID 6162 (GBS236) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 5; MW 21.6kDa).

Purified GBS236-GST is shown in Figure 208 (lane 6) and in Figure 225 (lanes 4-5).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1996

A DNA sequence (GBSx2106) was identified in *S.agalactiae* <SEQ ID 6165> which encodes the amino acid sequence <SEQ ID 6166>. This protein is predicted to be pantothenate metabolism flavoprotein homolog (dfp). Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2325(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9835> which encodes amino acid sequence <SEQ ID 9836> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG39941 GB:AF301375 MTW1216 [Methanothermobacter wolfeii
           prophage psiM100]
 Identities = 71/229 (31%), Positives = 117/229 (51%), Gaps = 27/229 (11%)

Query: 6   MKILITSGGTTEKIDTVRSITNHATGTLGKIIAEKYLREGHQVTLVTTKNAVKPESATNL 65
           +++L++ GGT E ID VR ITN ++G +G  +A +   +G  VTLV     V  +  + L
Sbjct: 172 LRVLVSLGGTLEPIDPVRVITNRSSGRMGLAVAREAYIQGADVTLVA--GTVSVDIPSQL 229

Query: 66  STFEIEDVDSLIKTLKPLVKEHDILIHSMAVSDYTPVYMADFEKVKSSDHLDTFLRKDNH 125
            T   E    + + +  L+ EHD+ + + AVSD+ PVY
Sbjct: 230 RTVRAETAHEMAEAVAELIGEHDVFVSAAAVSDFRPVYS--------------------- 268

Query: 126 EGKISSESEYQVLFLKKTPKVISLVKKWNPQITLVGFKLLVNVTKENLFKVARHSLIKNK 185
           E KISS+SE   L LK  PK+I + ++ NP+  +VGFK    V++E L   AR  +  +
Sbjct: 269 EEKISSDSEI-TLRLKPNPKIIRMARETNPEAFIVGFKAEHGVSEEELIAAARKQIEDSV 327

Query: 186 ATFILANDL-IDITSKHHIAYLLDHDNVYKATT--KEDIAQLIYEKVKK 231
           A  ++AND+ ++     +   ++  + V +  T  KE++A LI  ++ K
Sbjct: 328 ADMVVANDVSVEGFGSENNRAIIVSEGVTELPTMKKEELAGLIIGEIMK 376
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6167> which encodes the amino acid sequence <SEQ ID 6168>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1737(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 142/230 (61%), Positives = 170/230 (73%)

Query: 4    MAMKILITSGGTTEKIDTVRSITNHATGTLGKIIAEKYLREGHQVTLVTTKNAVKPESAT 63
            M MK++ITSGGTTE ID VR ITNH+TG LGK+I E++L+  H VTLVTTK A KP
Sbjct: 1    MTMKLIITSGGTTEPIDAVRGITNHSTGQLGKLITERFLQYHHDVTLVTTKTATKPLPNK 60
```

```
Query: 64  NLSTFEIEDVDSLIKTLKPLVKEHDILIHSMAVSDYTPVYMADFEKVKSSDHLDTFLRKD 123
            L   E+E V+ L+  LK  V  HDILIHSMAVSDYTPVYM D E+V  +D+L+ FL +
Sbjct: 61  RLRIIEVETVNDLMAALKDQVPHHDILIHSMAVSDYTPVYMTDLEQVSQADNLNCFLCEH 120

Query: 124 NHEGKISSESEYQVLFLKKTPKVISLVKKWNPQITLVGFKLLVNVTKENLFKVARHSLIK 183
           N E KISS S+YQVLFLKKTPKVIS VK+WNP I LVGFKLLVNV +E L KVAR SL K
Sbjct: 121 NSEPKISSASDYQVLFLKKTPKVISYVKQWNPNIKLVGFKLLVNVPQEELIKVARASLAK 180

Query: 184 NKATFILANDLIDITSKHHIAYLLDHDNVYKATTKEDIAQLIYEKVKKYD 233
           N A +ILANDL+DI +  H A L+ ++ V  A TKE IA L+YE++ K+D
Sbjct: 181 NHADYILANDLVDIQTGMHKALLISNNEVASADTKEAIADLLYERMTKHD 230
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1997

A DNA sequence (GBSx2107) was identified in *S.agalactiae* <SEQ ID 6169> which encodes the amino acid sequence <SEQ ID 6170>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.22   Transmembrane  117 - 133 ( 117 - 133)

----- Final Results -----
               bacterial membrane --- Certainty=0.1086(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9833> which encodes amino acid sequence <SEQ ID 9834> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07541 GB:AP001520 unknown conserved protein in B. subtilis
           [Bacillus halodurans]
 Identities = 94/221 (42%), Positives = 133/221 (59%), Gaps = 2/221 (0%)

Query: 52  AEKPFIWTEVFLREINRSNQEIILHIWPMTKTVILGMLDRELPHLELAKKEIISRGYEPV 111
           A + F + +     I +S     L  W    TV+LG+ D  LP ++   + +    ++ +
Sbjct: 27  ALQSFAYDDTLCTSIGKSQSPPTLRAWVHHNTVVLGIQDSRLPQIKAGIEALKGFQHDVI 86

Query: 112 VRNFGGLAVVADEGILNFSLVIPDVFERKLSISDGYLIMVDFIRSIFSDFYQPIEHFEVE 171
           VRN GGLAVV D GILN SLV+ +  E+  SI DGY +M + I S+F D  + IE  E+
Sbjct: 87  VRNSGGLAVVLDSGILNLSLVLKE--EKGFSIDDGYELMYELICSMFQDHREQIEAREIV 144

Query: 172 TSYCPGKFDLSINGKKFAGLAQRRIKNGIAVSIYLSVCGDQKGRSQMISDFYKIGLGDTG 231
            SYCPG +DLSI+GKKFAG++QRRI+ G+AV IYL V G    R++MI  FY   +
Sbjct: 145 GSYCPGSYDLSIDGKKFAGISQRRIRGGVAVQIYLCVSGSGAERAKMIRTFYDKAVAGQP 204

Query: 232 SPIAYPNVDPEIMANLSDLLDCPMTVEDVIDRMLISLKQVG 272
           +   YP + PE MA+LS+LL  P  V DV+ + L++L+Q G
Sbjct: 205 TKFVYPRIKPETMASLSELLGQPHNVSDVLLKALMTLQQHG 245
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6171> which encodes the amino acid sequence <SEQ ID 6172>. Analysis of this protein sequence reveals the following:

```
     Possible site: 61

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -0.22   Transmembrane   95 - 111 (  95 - 111)

----- Final Results -----
               bacterial membrane --- Certainty=0.1086(Affirmative) < succ>
```

```
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB07541 GB:AP001520 unknown conserved protein in B. subtilis
           [Bacillus halodurans]
 Identities = 97/228 (42%), Positives = 138/228 (59%), Gaps = 2/228 (0%)

Query: 30  ALSPFVWTEVFLKTINQEPNQLILHIWPMTRTVILGMLDRQLPYFELAKTEIGNNGYVPV 89
            AL  F + +    +I +  +   L  W    TV+LG+ D +LP  +     +    +  +
Sbjct: 27  ALQSFAYDDTLCTSIGKSQSPPTLRAWVHHNTVVLGIQDSRLPQIKAGIEALKGFQHDVI 86

Query: 90  TRNIGGLAVVADDGILNFSLVIPDHFSESISISNAYLIMVDVIRESFSDYYQRIEYHEIK 149
             RN GGLAVV D GILN SLV+ +   +  SI + Y +M ++I   F D+ ++IE  EI
Sbjct: 87  VRNSGGLAVVLDSGILNLSLVLKEE--KGFSIDDGYELMYELICSMFQDHREQIEAREIV 144

Query: 150 NSYCPGNFDLSIAGRKFAGIAQRRIKKGIVVSIYLSVCGDQAARGQLIKDFYEAGTQGEV 209
             SYCPG++DLSI G+KFAGI+QRRI+ G+ V IYL V G  A R ++I+ FY+    G+
Sbjct: 145 GSYCPGSYDLSIDGKKFAGISQRRIRGGVAVQIYLCVSGSGAERAKMIRTFYDKAVAGQP 204

Query: 210 TKVNYPQIDPECMATLSELLETPFTVAEVLERLRLTLRQLGFSLTEKS 257
            TK  YP+I PE MA+LSELL  P  V++VL +  +TL+Q G SL  +S
Sbjct: 205 TKFVYPRIKPETMASLSELLGQPHNVSDVLLKALMTLQQHGASLLTES 252
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 155/275 (56%), Positives = 199/275 (72%), Gaps = 8/275 (2%)

Query: 32  QDLAQLPVSIFKDYVTDAQDAEKPFIWTEVFLREINRSNQEIILHIWPMTKTVILGMLDR 91
            +DLA LP+ ++ D       A  PF+WTEVFL+ IN+   ++ILHIWPMT+TVILGMLDR
Sbjct: 10  RDLASLPIFVYGDGNKKVPGALSPFVWTEVFLKTINQEPNQLILHIWPMTRTVILGMLDR 69

Query: 92  ELPHLELAKKEIISRGYEPVVRNFGGLAVVADEGILNFSLVIPDVFERKLSISDGYLIMV 151
            +LP+ ELAK EI + GY PV RN GGLAVVAD+GILNFSLVIPD F   +SIS+ YLIMV
Sbjct: 70  QLPYFELAKTEIGNNGYVPVTRNIGGLAVVADDGILNFSLVIPDHFSESISISNAYLIMV 129

Query: 152 DFIRSIFSDFYQPIEHFEVETSYCPGKFDLSINGKKFAGLAQRRIKNGIAVSIYLSVCGD 211
            D IR  FSD+YQ IE+ E++ SYCPG FDLSI G+KFAG+AQRRIK GI VSIYLSVCGD
Sbjct: 130 DVIRESFSDYYQRIEYHEIKNSYCPGNFDLSIAGRKFAGIAQRRIKKGIVVSIYLSVCGD 189

Query: 212 QKGRSQMISDFYKIGLGDTGSPIAYPNVDPEIMANLSDLLDCPMTVEDVIDRMLISLKQV 271
            Q  R Q+I DFY+ G      + + YP +DPE MA LS+LL+ P TV +V++R+ ++L+Q+
Sbjct: 190 QAARGQLIKDFYEAGTQGEVTKVNYPQIDPECMATLSELLETPFTVAEVLERLRLTLRQL 249

Query: 272 GFN------DRLLMIRPDLVAEFNRFQAKSMANKG 300
            GF+      D+ L+   D V  + R Q + +  +G
Sbjct: 250 GFSLTEKSPDQALLTNFDAV--YERMQLEVVRKEG 282
```

A related GBS gene <SEQ ID 8951> and protein <SEQ ID 8952> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:    -16.85
GvH: Signal Score (-7.5): -5.07
     Possible site: 49
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value:  -0.22 threshold:  0.0
   INTEGRAL    Likelihood = -0.22   Transmembrane  117 - 133 ( 117 - 133)
   PERIPHERAL  Likelihood =  0.47      73
 modified ALOM score:   0.54

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.1086(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

```
                    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01564(451 - 1116 of 1518)
EGAD|13388|BS3758(27 - 249 of 281) hypothetical 31.4 kd protein in pta 3'region {Bacillus
subtilis} OMNI|NT01BS4391 hypothetical protein SP|P39648|YWFL_BACSU HYPOTHETICAL 31.4 KDA
PROTEIN IN PTA 3'REGION. GP|414014|emb|CAA51646.1||X73124 ipa-90d {Bacillus subtilis}
GP|2636300|emb|CAB15791.1||Z99123 alternate gene name: ipa-90d {Bacillus subtilis}
PIR|S39745|S39745 ywfL protein - Bacillus subtilis
%Match = 15.8
%Identity = 40.8  %Similarity = 61.0
Matches = 91  Mismatches = 82  Conservative Sub.s = 45

321       351       381       411       441       471       501       531
*WNLRETYWKISSDCDKINLAEFSRERMSDLLEWQDLAQLPVSIFKDYVTDAQDAEKPFIWTEVFLREINRSNQEIILHI
                                                     ||:: |   : :   : :
                           MANQPIDLLMQPKWRVIDQSSLGPLFDAKQSFAMDDTLCMSVGKGVSPATARS
                                   10        20        30        40        50

561       591       621       651       681       711       738       768
WPMTKTVILGMLDRELPHLELAKKEIISRGYEPVVRNFGGLAVVADEGILNFSLVIPDVFERK-LSISDGYLIMVDFIRS
|    |::||: |  || |:     : | ||  :||| |||||| |:|:|| ||:  |  |:| : |  ||  ||:::|
WVHHDTIVLGIQDTRLPFLQDGISLLESEGYRVIVRNSGGLAVVLDDGVLNISLIFED--EKKGIDIDKGYEAMVELMRR
               70        80        90       100       110         120       130

798       828       858       888       918                 972       996
IFSDFYQPIEHFEVETSYCPGKFDLSINGKKFAGLAQRRIKNGIAVSIYLSVCGDQKG--RSQMISDFYKIGLGD--TGS
::  :   || :|:| ||||| :|||||||||||::|||:: |:|| |||  | |: |  |: :|  ||:  | |
MLRPYNAKIEAYEIEGSYCPGSYDLSINGKKFAGISQRRVRGGVAVQIYL--CADKSGSERADLIRRFYQAALKDKQNDK
                150       160       170       180         190       200

1026      1056      1086      1116      1146      1176      1206      1236
PIAYPNVDPEIMANLSDLLDCPMTVEDVIDRMLISLKQVGFNDRLLMIRPDLVAEFNRFQAKSMANKGMVSRDE*CPR*F
   || : || ||:||:||   ::|:|::  :|  || :
KGVYPEIRPETMASLSELLQKDISVQDLMFALLTELKALSTHLYSAGLSIDEEMEFEKNLVRMAERNAKVFG
        220       230       240       250       260       270       280
```

SEQ ID 8952 (GBS390) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 73 (lane 7; MW 37kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 82 (lane 3; MW 62kDa).

GBS390-GST was purified as shown in Figure 216, lane 12.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1998

A DNA sequence (GBSx2108) was identified in *S.agalactiae* <SEQ ID 6173> which encodes the amino acid sequence <SEQ ID 6174>. This protein is predicted to be probable trimethylamine dehydrogenase (nemA). Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2218(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA83700 GB:Z33015 similar to trimethylamine DH [Mycoplasma
```

```
           capricolum]
 Identities = 162/311 (52%), Positives = 219/311 (70%), Gaps = 1/311 (0%)

Query: 3   NVQGNLFRPLTLPNGLSLENRFVLSPMVTNSSTSEGFVTDDDIAYAVRRAKSAPLQITGA 62
           N    LF P  L NG  LENRFVLSPM  + +T +G +TD +  Y  RR+ SAPLQITG
Sbjct: 2   NKYEKLFEPFYL-NGFKLENRFVLSPMTLSLATLDGKITDKEADYVKRRSHSAPLQITGG 60

Query: 63  AYITEYGQLFEYGFSVSKDEDIPGLTKLAKAMKSKGAKAVLQLTHAGRFSSHTLARHGYV 122
            Y  E+GQLFEYG S   D+DIP LT+L + MK+     +LQL HAG+FS  +L ++GY+
Sbjct: 61  VYFDEFGQLFEYGISAKSDDDIPSLTRLYQEMKTDSNCVILQLAHAGKFSKTSLKKYGYL 120

Query: 123 YGPSPMQLQSPYPHQVKELTHKDILRIIDEYVQATRRAIQAGFDGVEISSAQRLLIQTFF 182
           YGPS  +  +P  H+V EL  + I +II +Y  AT R I+AGF+G+EIS AQRLLIQTFF
Sbjct: 121 YGPSYEKNHTPIEHEVLELPKEKIKQIIQDYKDATLRVIKAGFNGIEISMAQRLLIQTFF 180

Query: 183 STFSNQRKDEYGPQTLTNRCRLGLEVFKAVQKVIREEAESDFILGFRATPEETRGSQIGY 242
           S   N+R DEY      NR R  LEV KA+++VI + A  +FI GFRATPEET G  +GY
Sbjct: 181 SQIINKRTDEYSATNFENRSRFCLEVVKAIREVIDKYAPKNFIFGFRATPEETYGDILGY 240

Query: 243 SIEEFMEFLEKILAIAQVDYLAIASWGHDVFRNTIRSEGVYKGQLVNQVIFEHFGDRVPI 302
           +IE+F++ ++KI+ I ++ YLAIASWGHD++ N +RS   YKGQLVN+VI++ + +++PI
Sbjct: 241 TIEDFIQLVDKIIEIGKISYLAIASWGHDIYLNKVRSNTKYKGQLVNKVIYDIYKNKLPI 300

Query: 303 MATGGINSASK 313
           +++GGIN+ +K
Sbjct: 301 ISSGGINTPTK 311
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6175> which encodes the amino acid sequence <SEQ ID 6176>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3055(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 265/390 (67%), Positives = 321/390 (81%)

Query: 8   LFRPLTLPNGLSLENRFVLSPMVTNSSTSEGFVTDDDIAYAVRRAKSAPLQITGAAYITE 67
           LF PLTLPNG  L+NRFVLSPMVTNSST +G+VT DD++YA+RRA SAPLQITGAAY+
Sbjct: 8   LFEPLTLPNGSQLDNRFVLSPMVTNSSTKDGYVTQDDVSYALRRAASAPLQITGAAYVDP 67

Query: 68  YGQLFEYGFSVSKDEDIPGLTKLAKAMKSKGAKAVLQLTHAGRFSSHTLARHGYVYGPSP 127
           YGQLFEYGFSV+KD DI GL +LA+AMK+KGAKAVLQLTHAGRF+SH L ++G+VYGPS
Sbjct: 68  YGQLFEYGFSVTKDADISGLKELAQAMKAKGAKAVLQLTHAGRFASHALTKYGFVYGPSY 127

Query: 128 MQLQSPYPHQVKELTHKDILRIIDEYVQATRRAIQAGFDGVEISSAQRLLIQTFFSTFSN 187
           MQL+SP PH+VK LT + I  +I  Y QATRRAIQAGFDGVE+SSAQRLLIQTFFSTFSN
Sbjct: 128 MQLRSPQPHEVKPLTGQQIEELIAAYAQATRRAIQAGFDGVEVSSAQRLLIQTFFSTFSN 187

Query: 188 QRKDEYGPQTLTNRCRLGLEVFKAVQKVIREEAESDFILGFRATPEETRGSQIGYSIEEF 247
           +R D YG QTL NR +L L V +AVQ+VI++EA   FI GFRATPEETRG+ IGYSI+EF
Sbjct: 188 KRTDSYGCQTLFNRSKLTLAVLQAVQQVIKQEAPDGFIFGFRATPEETRGNDIGYSIDEF 247

Query: 248 MEFLEKILAIAQVDYLAIASWGHDVFRNTIRSEGVYKGQLVNQVIFEHFGDRVPIMATGG 307
           ++ ++ +L +A++DYLAIASWG  VFRNT+RS G Y G+ VNQV+ ++  +++P+MATGG
Sbjct: 248 LQLMDWVLNVAKLDYLAIASWGRHVFRNTVRSPGPYYGRRVNQVVRDYLRNKLPVMATGG 307

Query: 308 INSASKVFEALQHAHMIGASTPLVVDPEFLQKIKAKCSDQINLRIKVSDLEGLAIPKASF 367
           +N+  K  EAL HA  IG STP VVDPEF  KIK  C + I+LRI+ +DL+ LAIP+ASF
Sbjct: 308 MNTPDKAIEALAHADFIGVSTPFVVDPEFAHKIKEGCEESIHLRIRPADLKSLAIPQASF 367

Query: 368 KDIVPLMDYGESLPKEAREVFRELRSNYRE 397
```

```
           KDIVPLMDYGESLPKE+R +FR L  NY+E
Sbjct: 368 KDIVPLMDYGESLPKESRTLFRSLTHNYKE 397
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1999

A DNA sequence (GBSx2109) was identified in *S.agalactiae* <SEQ ID 6177> which encodes the amino acid sequence <SEQ ID 6178>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3748(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB04594 GB:AP001510 unknown conserved protein [Bacillus halodurans]
 Identities = 121/333 (36%), Positives = 192/333 (57%), Gaps = 12/333 (3%)

Query: 1   MKLSVLDYGLIDYGKTASDAIQETILLSQEAERLGYHQFWVAEHHGVKAFSISNPELMIM 60
           MKLSVLD   I YG  A +A+++T  L++  E LGYH+FWV+EHH     + S+PE++I
Sbjct: 1   MKLSVLDQSPIAYGSNAKEALRQTTELAKVTEALGYHRFWVSEHHDASTLAGSSPEVLIA 60

Query: 61  HLANQTKSIKIGSGGIMPLHYSSFKLAETLKTLETCHPNRVSIGLGNSLGTVKVSNALRS 120
           HLA  TK I++GSGG+M  HYS++K+AE  K LE  HP R+ +GLG + G + ++
Sbjct: 61  HLAAHTKKIRLGSGGVMLPHYSAYKVAENFKLLEALHPGRIDVGLGRAPGGMPIAKMALQ 120

Query: 121 LHK---AHDYEEVLEELKSWLIDESSSKEPL----VQPTLSSFPDLYVLGSGQKSAYLAA 173
             K    H Y   ++++  +L D+  +          P + + PD+++LGS   SA +AA
Sbjct: 121 EGKEQNIHKYPLQVKDVIGYLQDDLPTDHRFHGLKATPLIDTVPDVWLLGSSGGSANVAA 180

Query: 174 KLGLGFTFGVFPFMDKDPLTEAKKLSSLYYHQFEEYYPNKSPNLMVAAFVVIADTSEEAE 233
           + G GF F    F++ +   +A +    Y   F+       P   VA FV+ ADT E+A+
Sbjct: 181 ENGTGFAFA--HFINGEGGVQAVE---SYRETFQPSALFDRPQTSVAIFVICADTDEQAD 235

Query: 234 NIAKTLDIWMLGNKDFNEFATFPTIEEANHYQLTPEQKAKIKSNRDRMIVGDPKQVKESL 293
            IA +LD+ ++  ++       P+IE A  Y  +P ++A+I+ NR RMIVG PK V++ L
Sbjct: 236 QIASSLDLSLIMLENGQLSKGTPSIESALSYPYSPFERARIRENRKRMIVGSPKAVRQQL 295

Query: 294 DALVNASQAEELLLIPLVPGLDQRIKSLKLLSQ 326
             L  A + EE++++ +    + RI+S +LL +
Sbjct: 296 VELARAYETEEVIVVTITHRFEDRIRSYELLGE 328
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6179> which encodes the amino acid sequence <SEQ ID 6180>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.60    Transmembrane  212 - 228 ( 210 - 229)

----- Final Results -----
               bacterial membrane --- Certainty=0.2041(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 173/329 (52%), Positives = 241/329 (72%), Gaps = 1/329 (0%)

Query: 1   MKLSVLDYGLIDYGKTASDAIQETILLSQEAERLGYHQFWVAEHHGVKAFSISNPELMIM 60
```

```
           MK+S+LDYG+ID  KT  +A+ ET  L+Q A++LG+H+FWVAEHH + AF+IS+PEL++M
Sbjct: 1   MKVSILDYGVIDKEKTPQEALLETRCLAQVADKLGFHRFWVAEHHNIYAFAISSPELLMM 60

Query: 61  HLANQTKSIKIGSGGIMPLHYSSFKLAETLKTLETCHPNRVSIGLGNSLGTVKVSNALRS 120
           HLA+ TK I+IGSGGIMPLHYSSFK+AE + TLE  HPNR+ +G+GNSLGT  V  AL S
Sbjct: 61  HLADHTKQIRIGSGGIMPLHYSSFKIAEWIMTLEALHPNRIDLGIGNSLGTTLVQRALSS 120

Query: 121 LHKAHDYEEVLEELKSWLIDESSSKEPL-VQPTLSSFPDLYVLGSGQKSAYLAAKLGLGF 179
           +H    Y +V+ EL  +L  +  S  P+ V P  +++P ++ L +  ++A LA +LGLG+
Sbjct: 121 IHCKDSYSQVVTELYQYLNPDHLSPLPIFVNPRGNTYPQIWTLSNSLETAELAGQLGLGY 180

Query: 180 TFGVFPFMDKDPLTEAKKLSSLYYHQFEEYYPNKSPNLMVAAFVVIADTSEEAENIAKTL 239
           TFG+FP++ KDP+TEAK++S+ Y   F      K P L++A F+V++DT E+AE +AK L
Sbjct: 181 TFGIFPYIPKDPITEAKRVSAHYRKAFRPSKLLKIPKLILAVFIVLSDTDEKAEALAKPL 240

Query: 240 DIWMLGNKDFNEFATFPTIEEANHYQLTPEQKAKIKSNRDRMIVGDPKQVKESLDALVNA 299
           DIWMLG +DFNEF T+P +EEA +Y LT +Q+  I +NR RM++G P  VK+ LD L+ A
Sbjct: 241 DIWMLGQQDFNEFKTYPDVEEARNYHLTEKQREAIAANRSRMVIGSPHTVKKQLDRLIEA 300

Query: 300 SQAEELLLIPLVPGLDQRIKSLKLLSQLY 328
            QA+ELL IPLVP    R ++L+LL+ LY
Sbjct: 301 CQADELLAIPLVPEFANRQRTLELLADLY 329
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2000

A DNA sequence (GBSx2110) was identified in *S.agalactiae* <SEQ ID 6181> which encodes the amino acid sequence <SEQ ID 6182>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2384(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF81345 GB:AC007767 Identical to a glycine cleavage system
           H-protein precursor from Arabidopsis thaliana gb|P25855.
           It contains a glycine cleavage H-protein domain
           PF|01597. ESTs gb|R90208, gb|AI
 Identities = 30/91 (32%), Positives = 53/91 (57%), Gaps = 1/91 (1%)

Query: 18  TISLTPELQDDLGTVGYVEFTD-DANLEVDDVILNIEASKTVMAILSPLTGKVVKVNTAA 76
           TI +T   QD LG V +VE  + ++++   +    +E+ K    ILSP++G+V++VNT
Sbjct: 59  TIGITDHAQDHLGEVVFVELPEANSSVSKEKSFGAVESVKATSEILSPISGEVIEVNTKL 118

Query: 77  SQEPTLLNSEKADENWLVVLTEVDYAAFEAL 107
           ++ P L+NS   ++ W++ +     A  EAL
Sbjct: 119 TESPGLINSSPYEDGWMIKVKPSSPAELEAL 149
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6183> which encodes the amino acid sequence <SEQ ID 6184>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3544(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 80/110 (72%), Positives = 98/110 (88%)

Query: 1   MKKIANYLLIEKNEELYTISLTPELQDDLGTVGYVEFTDDANLEVDDVILNIEASKTVMA 60
           MKKIANYLLIEK ++ YTIS+TPELQDD+GT+GY EFTD+ +L VDD+ILN+EASKTVM+
Sbjct: 1   MKKIANYLLIEKTDDRYTISMTPELQDDIGTIGYAEFTDNDHLAVDDIILNLEASKTVMS 60

Query: 61  ILSPLTGKVVKVNTAASQEPTLLNSEKADENWLVVLTEVDYAAFEALENA 110
           +LSPL G VV+ N AA+  PTLLNSEKA+ENW+VVLT+VD AAF+ALE+A
Sbjct: 61  VLSPLAGAVVERNEAATLTPTLLNSEKAEENWIVVLTDVDQAAFDALEDA 110
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2001

A DNA sequence (GBSx2111) was identified in *S.agalactiae* <SEQ ID 6185> which encodes the amino acid sequence <SEQ ID 6186>. This protein is predicted to be LRP16 (b1045). Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0608(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF15294 GB:AF202922 LRP16 [Homo sapiens]
 Identities = 73/171 (42%), Positives = 98/171 (56%), Gaps = 13/171 (7%)

Query: 88  DICLLQVDAIVNAANSKLLGCFIPNHHCIDNQIHTFAGSRLRLACHQLMTQQGRMEAVGQ 147
           DI  L+VDAIVNAANS LLG        +D  IH  AG  L   C  L + +      G+
Sbjct: 78  DITKLEVDAIVNAANSSLLG-----GGGVDGCIHRAAGPLLTDECRTLQSCK-----TGK 127

Query: 148 AKLTESYHLPCKYVIHTVGPYVKVDQKPSRIREDLLKSSYKSCLQLAVRANLKTIVFPCI 207
           AK+T  Y LP KYVIHTVGP    +   S+  E  L+S Y S L L +   L+++ FPCI
Sbjct: 128 AKITGGYRLPAKYVIHTVGPIAYGEPSASQAAE--LRSCYLSSLDLLLEHRLRSVAFPCI 185

Query: 208 STGEFGFPNQRAAELAVQAILEWQRENQHKL-YIIFNTFTPKDQDIYQKLL 257
           STG FG+P + AAE+ +  + EW  +++ K+  +I   F  KD+DIY+  L
Sbjct: 186 STGVFGYPCEAAAEIVLATLREWLEQHKDKVDRLIICVFLEKDEDIYRSRL 236
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6187> which encodes the amino acid sequence <SEQ ID 6188>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1992(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 139/266 (52%), Positives = 178/266 (66%), Gaps = 6/266 (2%)

Query: 1   MPNQKQLLLAMIEYLQSEKLTDVDDL----RTTDLQTVWRGLVNQQDPQNISQEYLSLED 56
           MP+   LL  MI  LQ+E+LT           T  Q +WR L+NQ+    +S++YL+LED
```

```
Sbjct: 1   MPSSFDLLGEMIGLLQTEQLTSSWACPLPNALTKRQDLWRALINQRPALPLSKDYLNLED 60

Query: 57  RYLSHWWNTQKVKTIDVCHQTVYSNVFTYHGDICLLQVDAIVNAANSKLLGCFIPNHHCI 116
            YL  W  +    ++  C +T Y+++F YHGDI  L VDAIVNAANS+LLGCF PNH CI
Sbjct: 61  AYLDDWRASFVPVSVKDCQKTNYTSLFLYHGDIRYLAVDAIVNAANSELLGCFSPNHGCI 120

Query: 117 DNQIHTFAGSRLRLACHQLMTQQGRMEAVGQAKLTESYHLPCKYVIHTVGPYVKVDQKPS 176
           DN IHTFAGSRLRLAC  +MT+QGR EA+GQAKLT +YHLP  Y+IHTVGP +      S
Sbjct: 121 DNAIHTFAGSRLRLACQAIMTEQGRKEAIGQAKLTSAYHLPASYIIHTVGPRITKGHHVS 180

Query: 177 RIREDLLKSSYKSCLQLAVRANLKTIVFPCISTGEFGFPNQRAAELAVQAILEWQRENQH 236
            IR DLL   Y+S L LAV+A L ++ F  ISTGEFGFP + AA++A++ +L+WQ E+
Sbjct: 181 PIRADLLARCYRSSLDLAVKAGLTSLAFCSISTGEFGFPKKEAAQIAIKTVLKWQAEHPE 240

Query: 237 K--LYIIFNTFTPKDQDIYQKLLLKE 260
              L  IFNTFT +D+ +Y   L KE
Sbjct: 241 SKTLTTIFNTFTSEDKALYDTYLQKE 266
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2002

A DNA sequence (GBSx2112) was identified in *S.agalactiae* <SEQ ID 6189> which encodes the amino acid sequence <SEQ ID 6190>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2171(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6191> which encodes the amino acid sequence <SEQ ID 6192>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2477(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 218/284 (76%), Positives = 250/284 (87%)

Query: 4   WKTLEKTNHSQSEILSQLIEESDAIVVGIGAGMSAADGFTYIGPRFEEAFPDFIAKYQLL 63
           W T  + N +Q+E L+QLI+E+DA+VVGIGAGMSAADGFTYIG RFE AFPDFIAKYQ L
Sbjct: 4   WTTYPQKNLTQAEQLAQLIKEADALVVGIGAGMSAADGFTYIGSRFETAFPDFIAKYQFL 63

Query: 64  DMLQASLYDFEDWEEYWAFQSRFVALNYLDQPVGQAYLDLKDILAKKEYHIITTNADNAF 123
           DMLQASL+DFEDW+EYWAFQSRFVALNYLDQPVGQ+YLDLK+IL  K+YHIITTNADNAF
Sbjct: 64  DMLQASLFDFEDWQEYWAFQSRFVALNYLDQPVGQSYLDLKEILGTKDYHIITTNADNAF 123

Query: 124 AVADYNLEKVFHIQGEYGLWQCSQHCHQQTYRNDQAIRQMIAQQKDMKIPSNLIPKCPKC 183
            VA Y+   +FHIQGEYGLWQCSQHCHQQTY++D  IRQMIA+QK+MK+P  LIP CP+C
Sbjct: 124 WVAGYDPHNIFHIQGEYGLWQCSQHCHQQTYKDDTVIRQMIAEQKNMKVPGQLIPHCPEC 183

Query: 184 DQPFEINKRNEEKGMVEDADFHAQRQRYENFLSQHQNDKVLYLEIGVGHTTPQFIKHPFW 243
           + PFEINKRNEEKGMVEDADFHAQ+ RYE FLS+H+  KVLYLEIGVGHTTPQFIKHPFW
```

```
Sbjct: 184 EAPFEINKRNEEKGMVEDADFHAQKARYEAFLSEHKEGKVLYLEIGVGHTTPQFIKHPFW 243

Query: 244 RFVSLNENSLFVTLNHKHYRIPQKIRSRSVQLTQHIAELIAEAK 287
           + VS N N+LFVTLNHKHYRIP  IR +S++LT+HIA+LI+  K
Sbjct: 244 KRVSENPNALFVTLNHKHYRIPLSIRRQSLELTEHIAQLISATK 287
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 2003

A DNA sequence (GBSx2113) was identified in *S.agalactiae* <SEQ ID 6193> which encodes the amino acid sequence <SEQ ID 6194>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1086(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12865 GB:Z99109 similar to lipoate-protein ligase [Bacillus subtilis]
 Identities = 130/331 (39%), Positives = 206/331 (61%), Gaps = 5/331 (1%)

Query: 9   NGKRITDGAIALAMQVYILQNVFLDDDILFPYYCDPKVEIGKFQNAVIETNQEYLKEHDI 68
           + + I D  I LA++ Y ++++  +   L  Y   P + IGK QN + E N +Y++E+ I
Sbjct: 5   DNQNINDPRINLAIEEYCVKHLDPEQQYLLFYVNQPSIIIGKNQNTIEEINTKYVEENGI 64

Query: 69  PVVRRDTGGGAVYVDSGAVNICYLMKDHGQ-FGDFKRAYEPAIKALKTLGASSVEMRERN 127
            VVRR +GGGAVY D G +N  ++ KD G  F +FK+  EP I+AL  LG  + E+  RN
Sbjct: 65  IVVRRLSGGGAVYHDLGNLNFSFITKDDGDSFHNFKKFTEPVIQALHQLGVEA-ELSGRN 123

Query: 128 DLVIDGKKVSGAAMTIVNGRIYGGYSLLLDVDFDAMEKVLNPNRKKIESKGIKSVRSRVG 187
           D+V+DG+K+SG A     GRI+   +L+ D   D +   L   + KIESKGIKS+RSRV
Sbjct: 124 DIVVDGRKISGNAQFATKGRIFSHGTLMFDSAIDHVVSALKVKKDKIESKGIKSIRSRVA 183

Query: 188 DIRSHLSEDYRHITTDQFKDLMVCQLLHIDHIDQAKRYHLTEKDWAAIDALADEKYKNWD 247
           +I   L +    +TT++F+  ++  + + + +     Y LTEKDW  I  ++ E+Y+NWD
Sbjct: 184 NISEFLDDK---MTTEEFRSHLLRHIFNTNDVGNVPEYKLTEKDWETIHQISKERYQNWD 240

Query: 248 WNYGNSPQYSYHRDARFPSGTYDFHLEIEKGIITNCRIYGDFFSSKDISDIENLLIGCPM 307
           WNYG SP+++ +   R+P G+ D HLE++KG I +C+I+GDFF   D+S+IENLL+G
Sbjct: 241 WNYGRSPKFNLNHSKRYPVGSIDLHLEVKKGKIEDCKIFGDFFGVGDVSEIENLLVGKQY 300

Query: 308 KEELVLEKLSTLSLEDYFGQTSPEEIKAVLF 338
           +  ++ + L  ++L+ YFG  + E+   +++
Sbjct: 301 ERSVIADVLEGVNLKHYFGNITKEDFLDLIY 331
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 6195> which encodes the amino acid sequence <SEQ ID 6196>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0939(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 248/339 (73%), Positives = 283/339 (83%)
```